Functional Data Analysis with Applications in Finance

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftwissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
M.Sc. Michal Benko
geboren am 31.07.1979 in Banská Bystrica

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Wirtschaftwissenschaftlichen Fakultät:
Prof. Oliver Günther Ph.D.

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Prof. Dr. Alois Kneip

eingereicht am:	23. November 2006
Tag des Kolloquiums:	20. Dezember 2006

Abstract

In many different fields of applied statistics an object of interest is depending on some continuous parameter. Typical examples in finance are implied volatility functions, yield curves or risk-neutral densities. Due to the different market conventions and further technical reasons, these objects are observable only on a discrete grid, e.g. for a grid of strikes and maturities for which the trade has been settled at a given time-point. By collecting these functions for several time points (e.g. days) or for different underlyings, a bunch (sample) of functions is obtained – a *functional data set*. The first topic considered in this thesis concerns the strategies of recovering the functional objects (e.g. implied volatilities function) from the observed data based on the nonparametric smoothing methods. Besides the standard smoothing methods, a procedure based on a combination of nonparametric smoothing and the no-arbitrage-theory results is proposed for implied volatility smoothing. The second part of the thesis is devoted to the functional data analysis (FDA) and its connection to the problems present in the empirical analysis of the financial markets. The theoretical part of the thesis focuses on the functional principal components analysis – functional counterpart of the well known multivariate dimension-reduction-technique. A comprehensive overview of the existing methods is given, an estimation method based on the dual problem as well as the two-sample inference based on the functional principal component analysis are discussed. The FDA techniques are applied to the analysis of the implied volatility and yield curve dynamics. In addition, the implementation of the FDA techniques together with a FDA library for the statistical environment XploRe are presented.

Keywords:
Functional Data Analysis, Implied Volatility, Principal Component Analysis, Bootstrap

Zusammenfassung

An vielen verschiedenen Stellen der angewandten Statistik sind die zu untersuchenden Objekte abhängig von stetigen Parametern. Typische Beispiele in Finanzmarktapplikationen sind implizierte Volatilitäten, risikoneutrale Dichten oder Zinskurven. Aufgrund der Marktkonventionen sowie weiteren technisch bedingten Gründen sind diese Objekte nur an diskreten Punkten, wie zum Beispiel an Ausübungspreise und Maturitäten, für die ein Geschäft in einem bestimmten Zeitraum abgeschlossen wurde, beobachtbar. Ein funktionaler Datensatz ist dann vorhanden, wenn diese Funktionen für verschiedene Zeitpunkte (z.B. Tage) oder verschiedene zugrundeliegende Aktiva gesammelt werden. Das erste Thema, das in dieser Dissertation betrachtet wird, behandelt die nichtparametrischen Methoden der Schätzung dieser Objekte (wie z.B. implizierte Volatilitäten) aus den beobachteten Daten. Neben den bekannten Glättungsmethoden wird eine Prozedur für die Glättung der implizierten Volatilitäten vorgeschlagen, die auf einer Kombination von nichtparametrischer Glättung und den Ergebnissen der arbitragefreien Theorie basiert. Der zweite Teil der Dissertation ist der funktionalen Datenanalyse (FDA), speziell im Zusammenhang mit den Problemen, der empirischen Finanzmarktanalyse gewidmet. Der theoretische Teil der Arbeit konzentriert sich auf die funktionale Hauptkomponentenanalyse – das funktionale Ebenbild der bekannten Dimensionsreduktionstechnik. Ein umfangreicher Überblick der existierenden Methoden wird gegeben, eine Schätzmethode, die von der Lösung des dualen Problems motiviert ist und die Zwei-Stichproben-Inferenz basierend auf der funktionalen Hauptkomponentenanalyse werden behandelt. Die FDA-Techniken sind auf die Analyse der implizierten Volatilitäten- und Zinskurvendynamik angewandt worden. Darüber hinaus, wird die Implementation der FDA-Techniken zusammen mit einer FDA-Bibliothek für die statistische Software Xplore behandelt.

Schlagwörter:
Funktionale Datenanalyse, Implizierte Volatilität,
Hauptkomponentenanalyze, Bootstrap

Contents

List of Figures

List of Tables

Chapter 1

Introduction

After publishing the seminal pricing paper by Black and Scholes [1973] the derivative markets gained a lot of attention of the investors, since the financial derivatives like options give the market participants the opportunities to either trade their expectations about the future development (or states) of the market (speculations) or use them for protection against risky-events (hedges). The Black-Scholes (BS) pricing formula gives a simple formula for pricing. Empirical studies showed that the BS model is too simple to be able to explain all important phenomena observed in the financial markets. However, as argued by Black [1992]: "*Yet that weakness (simplicity) is also its greatest strength. People like the model because they can easily understand its assumptions. The model is often good as a first approximation, and if you can see the holes in the assumptions you can use the model in more sophisticated ways.*"

This statement announces the future destiny of the BS model: practitioners until now quote options in terms of the volatility implied by the BS model – implied volatility, knowing on the one hand that the BS model is inappropriate to be used for pricing directly, but are used to the BS model and interpretation of its parameter e.g. as a measure of nervousity (expected risk) of the market. On the other hand, further research proposed different generalizations of the BS model: jump diffusion model proposed by Merton [1976], stochastic volatility model by Heston [1993] or combination of the last two approaches proposed by Bates [1996]. Many other models have been considered, in general, using certain set of assumptions on the stock-markets and "pushing" them through the general arbitrage-free/martingale theory (fundamental theorem of asset pricing – FTAP) first proved by Harrison and Kreps [1979] – a theory that is widely accepted until today.

Renault [1997] argues, referring to Melino's lectures: "*...when one wants to speculate on the reasons for the widespread adoption of continuous time-*

models in asset pricing, one could argue that they have been adopted not because of their empirical properties but in spite of them". Renault [1997] is noting here that the FTAP based models are often used under "implausibly restrictive statistical specifications" and are still not able to mimic important empirical features observed in the financial markets.

This thesis follows these arguments and focuses, in the empirical part, on the static and dynamic empirical properties and phenomena of the option prices and related topics (in particular the state-price densities and yield curves that will be discussed in addition) and does not stack to specific pricing models. By analyzing the option prices, it seems to be advantageous to transform the option-price function. A popular transformation is the transformation to the already mentioned BS Implied Volatility – I would like to stress the fact that the implied volatility is not primary understood here as a parameter implied by the Black-Scholes model but as a transformation of the option-price with its own interpretation. Detailed discussion can be found in the chapter 2 and 3.

Option on a given asset and at a given time has two important parameters – exercise price (K) and time-to-maturity (τ). Consequently, the implied volatility (transformed option price) is also a function of K and τ, these can be in general treated as continuous parameters.

However, only options on some discrete grid of these parameters can be observed on the market in a given time. This is caused partially by the market mechanism and conventions and partially by the different trading intensity for different parameters (liquidity issues). For similar reasons, observed data can be contaminated by some (observational) noise. The first topic considered in this thesis concerns the strategies of recovering the implied volatilities function (surface) from the observed data. Since, as argued above, the strategy is to *let the data speak for them self*, nonparametric estimation techniques are natural candidates. For an overview of nonparametric methods see Härdle [1990] among others.

Neglecting, for a moment the necessity to recover the functions from a discrete data set, and collecting these functions for several time points (e.g. days) or for different underlyings a bunch (sample) of functions is obtained – a *functional data set*. The analysis of the functional data set is often referred to as Functional Data Analysis (FDA). A monograph on the FDA by Ramsay and Silverman [2005] summarizes the typical models considered in the FDA and most of the popular FDA techniques.

The questions arising by the statistical analysis of functional data are basically identical to the standard statistical analysis of univariate or multivariate objects. In fact, without knowing (or assuming) some additional information about the underlying functions, this will be the only appropriate

way to deal with the functional data set.

If, however, some "functional" feature (quality) of the underlying data-set can be assumed, then this information can be incorporated into the estimation procedure and improve the quality of the resulting estimate. A typical example of a functional quality is the "smoothness" of the underlying functions w.r.t. its parameters – smoothness is a nice example since it does not have direct meaning, if the data set is regarded as a generated by a random vector (in the multivariate approach) but can be addressed directly if the data is modeled as a collection of random functions (random variables with realizations in some proper functional space). Using the no-arbitrage arguments introduced by Breeden and Litzenberger [1978] the discounted second derivative of the call-price function (consequently the function of the first two derivatives of the IV function, see section B.1) is so called state-price-density function. Similarly the function of the first derivative of the yield curve has an economical interpretation as the forward yield curve. Hence, the FDA approach seems to be well motivated in the empirical finance.

Many different methods known from typical multivariate techniques, beginning with parametric (linear) regression models, (many of them, together with their applications are summarized in Ramsay and Silverman [2005] and Ramsay and Silverman [2002]) up to the nonparametric (nonparametric in functional sense) methods recently published in a monograph by Ferraty and Vieu [2006] have been introduced to the FDA.

The theoretical part of the thesis focuses on the Functional Principal Components Analysis (FPCA) – functional counterpart of the well known multivariate dimension-reduction-technique. A very nice example of the advantages of the functional approach is the Smoothed FPCA (SPCA), proposed by Silverman [1996] and studied by Pezzulli and Silverman [1993] that enables direct combination of the FPCA analysis together with a general smoothing approach that makes the use of the information stored in some linear differential operators possible. An important application of the FPCA already known from multivariate PCA, is motivated by the Karhunen-Loève decomposition of a random function to the set of functional parameters – factor functions and corresponding factor loadings (scalar random variables). This application is much more important than in the standard multivariate PCA since the distribution of the random function is in general too complex to be directly analyzed and the Karhunen-Loève decomposition reduces the analysis to the interpretation of the factor functions and the distribution of scalar random variables. This argument can also be used to construct the two-sample inference based on the FPCA method, proposed recently by Benko et al. [2006b].

The main goal of this thesis is to discuss the possibilities of the imple-

mentation of the nonparametric techniques and functional data analysis in the empirical analysis of financial markets. The thesis includes the discussion of the data sources needed for empirical analysis, implementation in the statistical software, discussion of the theoretical properties of the methods, as well as the applications of the methods to the German (European) financial markets, in particular the implied volatilities for the option of the German Stock Index (ODAX), as well as the EURIBOR interest rates.

The thesis is organized as follows:

Chapter 2 introduces the financial theory needed for the further chapters. Basic stochastic model of a financial market, and financial derivatives as futures, options and swaps are described. The arbitrage-free/martingale theory as well as the Black-Scholes market are introduced. At the end of the chapter, the concept of the Black-Scholes Implied Volatility is introduced and the interpretation of this concept is discussed.

Chapter 3 consists of two parts, first part – sections 3.1 to 3.4 – are devoted to nonparametric techniques, focused on the local regression and density estimation as well as the functional series estimators. Practical issues as well as the theoretical properties needed in further chapters are discussed. In the second part – sections 3.5 to 3.8 – the raw data (option data, interest rates), its properties and calculating procedure of the implied volatilities are discussed. The phenomena of IVs design (liquidity issues) and problem of estimating the IV function from the noisy data are discussed and the combination of nonparametric techniques with arbitrage-free theory (via the concept of state-price-density), originally introduced by Benko et al. [2006a] are discussed in section 3.8.

After introducing the basic setup of the functional data analysis and the implementation issues in the first section, chapter 4 consists again of two parts. In the first part, section 4.2 discusses the functional principal components analysis and smoothed (regularized) principal components. Two approaches are considered. First the approach based on the functional basis estimation as proposed by Ramsay and Silverman [2005]. The library for Functional Data Analysis developed for XploRe – statistical environment is described in addition in Appendix A. The described methods are applied to the EURIBOR yield curves. Selected aspects of the visualization of the FDA are considered. The first advantage compared to standard multivariate methods is that the FPCA model can be evaluated on arbitrary fine grid. Furthermore, besides the classical factors with well known interpretation as level, slope and curvature for longer maturities, a factor capturing the strong variation of the maturities shorter than 1 month has been identified. The other estimation technique for FPCA, proposed by Benko et al. [2006b] is based on the duality relations between row and column space of the data

matrix, well known from the multivariate PCA. The theoretical properties of the proposed method, the finite sample properties illustrated by the simulation study as well as implementation issues are discussed. In the second part of chapter 4 the two-sample inference based on the FPCA is discussed, a bootstrap based testing procedure is proposed and studied. The method is applied to the IV analysis, motivated by the fact that the estimated factor function for two different time-to-maturity (1M, 2M) groups are of similar structure. It has been shown that for both groups $L = 3$ components suffice to explain more than 95% of the variability of the returns of the implied volatility functions. An application of the tests developed in Section 4.3 does not reject the equality of the corresponding eigenspaces. Our analysis overcomes the limitations of the similar study by Fengler et al. [2003] based on discretized vectors of functional values by providing the tests in a fully functional setup. This part of the chapter is based on the work originally proposed in Benko et al. [2006b].

Major part of the computations were done in XploRe – statistical environment for statistical computing partially developed at the C.A.S.E., Humboldt-University and MD*Tech, limited.

Chapter 2

Options Markets Beyond Black-Scholes

In this chapter, the basic mathematical structure of financial market models is presented. The aim is not to give an exhausting study or a complete overview of the literature connected to this topic. The aim is to define all economic and probabilistic concepts that will be necessary for the later chapters and present them in a compact way, for deeper discussion see Musiela and Rutkowski [1997], Shreve [2004], and Dupačová et al. [2002], compare also with Hafner [2004], chapter 2 and Fengler [2005b], chapter 1.

Consider a financial market with a set of assets such as bonds or equities. Besides these *primary* assets also *secondary* instruments are often traded – instruments whose payments dependent on some primary (*underlying*) asset (or some other factors) – the *derivative* market. An example for a simple derivative is the *European call option*, that gives the buyer the right to buy an underlying asset for a predefined price K (*Strike* price) at some future time point (*expiry* date) T. A *European put option* gives the buyer the right to sell an underlying asset for a predefined strike. Alternatively European call and put can be defined as instruments that yield at T the payoff $\max(S_T - K, 0)$ or $\max(K - S_T, 0)$ respectively. These simple options are often called *plain vanilla*. Alternatively to European styled derivatives the *American styled* derivatives can be exercised at any time up to the expiry date T. Nowadays these options are standardized and frequently traded on financial markets. Clearly, nobody will give any right for free. The buyer has to pay an *option price* to the seller. The financial mathematics develops, based on certain economical considerations, pricing methods and pricing models for determining the option price. Although, from purely rational argumentation the price of an option is some agreement between seller and buyer and the price should be determined by proper supply-demand or behavioral considerations, it was

especially the development of the pricing models such as Black and Scholes [1973] and their generalizations that in fact motivated the boom of the derivative markets. The main aim of this short introduction is to give the basic formal ideas that enable the pricing of financial derivatives.

Of course a pricing model has to be based on a proper model of the financial market:

First, consider a *continuous-time* financial market (model) – trades can take place continuously during some trading period $t \in [0, T_M]$, $T_M > 0$. A *frictionless* market is assumed, i.e. no "transaction costs", "no taxes", no conditions on "short sales", infinitesimally divisible assets and, if not written otherwise, "no-dividends". For our further consideration only value (price)-development over time will be important, hence only this has to be modeled. Since typically the value-development of an asset is not known in advance – it is uncertain, one can model it as a real valued stochastic process. The uncertain world of a financial market is described by a probability space $(\Omega, \mathcal{F}, P)$ where Ω is a set of possible elementary states, $\mathcal{F}$ is a sigma algebra on Ω representing the information structure on the market and P is a probability measure – the *objective probability measure*. As, in real life investors can use only information available up to the current time t a filtration $\mathcal{F}_t, t = [0, T_M]$ ($\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$) is defined, where $\mathcal{F}_t$ stands for the set of information available to the investor at the time t. Assume that $\mathcal{F}_t$ is right continuous and due to technical reasons $\mathcal{F}_0$ contains all P-null sets, without lost of generality assume $\mathcal{F} = \mathcal{F}_{T_M}$.

Consider a market with $d + 1$ assets and denote their price processes by $S_t = (S_{t,0}, S_{t,1}, \ldots, S_{t,d})^\top$, for all $t \in [0, T_M]$. Clearly, price-processes $S_{t,i}, i = 1, \ldots, d$ should be at least $\mathcal{F}_t$-*adapted*, i.e. $\sigma(S_{s,i}, s \leq t) \subset \mathcal{F}_t$ for $t \in [0, T_M]$. Furthermore the existence of some type of *saving account* is assumed, where the investment cannot "fully disappear", let us call this asset *numéraire*. Assume that the 0-th asset is the numéraire, hence assume that the $S_0 = \{S_{t,0},\ t \in [0, T_M]\}$ is P-a.s. positive. The so-called *discounted* price processes are denoted by $S_t^* = (\frac{S_{t,i}}{S_{t,0}}, i = 0, \ldots d,\ t \in [0, T_M])$.

The last important assumption that needs to be discussed is the *no-arbitrage* assumption. Roughly speaking an arbitrage-opportunity is a possibility for an investor to set-up a *strategy* that with positive probability "makes money from nothing". To be more formal, a definition of strategies is needed and the meaning of "making money from nothing" needs to be explained.

A trading strategy over some time-interval $[0, T]$ is an $\mathcal{F}_t$-predictable process (for our purposes $\mathcal{F}_t$-adapted and left continuous) $\theta = \{\theta_{t,i},\ i = 0, \ldots, d,\ t \in [0, T_M]\}$ such that the $\int_0^t \theta_s dS_s < \infty$ and $\int_0^t \theta_s dS_s^* < \infty$. The value $\theta_{i,t}$ is the number of units of the i-th asset in the portfolio (as infinites-

imally divisible assets and market with "short selling" are assumed, this can be any (finite) real value). The value of the trading strategy θ is defined by $V_t(\theta) = \theta^\top S_t$. Very interesting are strategies that require only an initial investment $V_0(\theta)$ and afterwards are self financing. Formally a trading strategy with value process V_t is called *self financing* if $dV_t(\theta) = \sum_{i=1}^{d} \int_0^t \theta_{s,i} dS_{s,i}$ for $t \in [0, T]$. The next technical constraint on the strategy is its lower bound, see Musiela and Rutkowski [1997], who call strategies with bounded discounted-value from below for each $t \in [0, T]$ *tames*. In the following text, all strategies are assumed to be tames. The motivation of this condition seems to be of technical nature, however as argued for example in Hafner [2004] "...there is usually a limit of how much loss an investor is willing to tolerate", hence the "tame"-condition has a natural interpretation.

Finally, a self-financing strategy is called an *arbitrage-opportunity* iff $V_0(\theta) = 0$ P-a.s., $V_T(\theta) \geq 0$ P-a.s. and $P(V_T(\theta) > 0) > 0$. In the following the financial market is considered to be *arbitrage-free* – without any arbitrage-opportunity. (In practice there are typically arbitrage-opportunities, however are usually immediately detected by some investor and corrected by the markets supply or demand.)

At this place it is convenient to show the connection between the no-arbitrage assumption and *equivalent martingale measures* . This connection is so important for pricing theory that it is usually refereed to as *Fundamental Theorem of Asset Pricing* (FTAP), and can be stated as:

A market is arbitrage-free iff there exists a measure Q defined on $(\Omega, \mathcal{F})$ equivalent to P (i.e. with same null-sets) such that the discounted price process S_t^* is a Q-martingale, i.e. $E_Q(S_t|\mathcal{F}_s) = S_s$ for $0 \leq s \leq t$.

Strictly speaking, the implication: if there exists some martingale measure the market is arbitrage-free is correct without further limitation, the reverse implication needs to be stated more carefully using the "no free lunch with wanishing risk"-condition, see Delbaen and Schachermayer [1994]. An easy readable proof of this theorem, however for the discrete-time case can be found in Föllmer and Schied [2002].

Unfortunately, in general the equivalent martingale measure Q is not unique. The set of equivalent martingale measures will be denoted by $\mathcal{Q}$. The next step on the way to pricing financial derivatives, is naturally to define a derivative. Until now, only the primary assets have been considered. A (European styled) *derivative* (or contingent claim) H with expiry date T is some nonnegative ($\mathcal{F}_T$-measurable) random variable such that $E_Q(H|S_{T,0}) < \infty$ for all $Q \in \mathcal{Q}$.

The random variable H models in fact the payoff of the derivative, in

general this payoff is determined by a function of the underlying primary asset (or assets) – the *payoff function* $H(S_t)$, $t = [0, T]$. Plain vanilla calls and puts mentioned above have, for example, the pay-off functions $\max(S_t - K; 0)$ and $\max(K - S_t, 0)$ respectively.

The basic idea of pricing based on the no-arbitrage consideration – is to find a proper (self-financing) strategy that gives the same payoff. Then the "price of the derivative" and the "value of the strategy" should coincide under no-arbitrage conditions. Unfortunately this is not always possible. The set of derivatives for which this is reasonably possible is called *attainable*.

A derivative H with expiration time T is called attainable if there exists an admissible strategy θ such that $V_T(\theta) = H$ P-a.s. A strategy θ is called admissible if it is self financing and if there exists some $Q^* \in \mathcal{Q}$ such that $V^*(\theta)$ is Q^* martingale. Such a strategy θ is called a *replicating strategy*.

Clearly, under no-arbitrage, the value process $V_t(\theta)$ and the price of the claim denoted by $\Pi_t(H)$ must coincide for $t \in [0, T]$. Furthermore it can be shown, see Hafner [2004] p. 19 that

$$\Pi_t(H) = E_{Q^*}\left(S_{T,0}^{-1} H | \mathcal{F}_t\right) \quad \text{for all } t \in [0, T]. \tag{2.1}$$

and that this price process is invariant to the choice of the equivalent martingale measure, see Musiela and Rutkowski [1997] p. 235.

It should be mentioned that the arbitrage-free pricing corresponds to the assumption that the investors acting on the market are *risk-neutral*. Hence, (2.1) is often referred to as the *risk-neutral pricing* formula, see Cox and Ross [1976]. For this reason the equivalent martingale measure is also often called *risk-neutral*. Since arbitrage-free (or risk-neutral) pricing enables to price the attainable derivatives only, two important types of markets are distinguished, first markets where each derivative with an arbitrary expiration date is attainable – *complete markets*. Other markets are called *incomplete*. It should be mentioned that the completeness of a market is equivalent to the uniqueness of the risk-neutral measure Q, see Föllmer and Schied [2002] for a nicely readable proof in discrete time, and hence the complete markets are in fact easy to handle.

In the next section a simple but famous financial model will be introduced – the Black-Scholes Market.

2.1 Black-Scholes Market

Black and Scholes [1973], partially motivated by Samuelson [1965] assumed that the evolution of the stock-price process S_t can be described by the

following stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{2.2}$$

where μ denotes a constant (instantaneous) drift (appreciation rate), σ is a constant (instantaneous) *volatility* and finally W_t denotes a standard Wiener process defined on $(\Omega, \mathcal{F}, P)$. A frictionless market is assumed. First part of this section follows partially the arguments presented in Musiela and Rutkowski [1997], section 5.1.

At this point it is worth to emphasize that (2.2) is only a symbolic notation for following integral equation:

$$S_t = S_0 + \int_0^t \mu S_u du + \int_0^t \sigma S_u dW_u, \ t \in [0, T_M]. \tag{2.3}$$

The information filtration $\mathcal{F}_t$ is assumed to be the P-augmented canonical filtration of W_t, i.e. $\mathcal{F}_t^W = \sigma(W_s, s \leq t)$, for some technical reasons each $\mathcal{F}_t$ needs to contain also the P-null sets of $\mathcal{F}$. Using Itô's formula, see Dupačová et al. [2002], one can check that the solution of (2.2) is given by:

$$S_t = S_0 e^{\sigma W_t + (\mu - 0.5\sigma^2)t}, \ t \in [0, T_M]. \tag{2.4}$$

As by (2.4) S_t is an invertible function of W_t it is clear that the filtrations of W_t and S_t coincide and hence in fact it is assumed that the information on the market is based purely on the observation of the stock-price process. The next important implication of (2.4) is that the stock returns are *log-normally* distributed, i.e. the so called log-return:

$$\ln(S_t/S_u) = \ln(S_t) - \ln(S_u)$$

is normally distributed (under the objective measure P) with mean $(\mu - 0.5\sigma^2)(t-u)$ and variance $\sigma^2(t-u)$ for $u \leq t \leq T_M$.

The next important (but standard) assumption is the existence of a *risk-free security* – a special type of numéraire mentioned in the previous section. The risk-free security can be described by the price process $B_t = e^{rt}$, $t \in [0, T_M]$ or by analogue of (2.2) by $dB_t = rB_t dt$ with the boundary condition $B_0 = 1$. The parameter r is called *risk-free interest rate* and is assumed to be constant over the trading period $[0, T_M]$. It is assumed that the borrowing and lending interest rates both equal r. It can easily be shown, see Hafner [2004] p. 24-25 that the Black-Scholes market with two securities, a risk-free security B_t and a risky asset S_t, is complete and hence the risk-neutral measure

Q is unique. In fact the completeness of the market and its transparent assumptions yield the popularity of the BS model.

Clearly the risk-neutral pricing formula (2.1) remains valid under this numéraire. Assuming the existence of the risk-free interest rate the price $\Phi(H)$ of a derivative H with expiration date T is

$$\Pi_t(H) = e^{-r(T-t)} E_Q\left(H|\mathcal{F}_t\right) \quad \text{for all } t \in [0,T]. \tag{2.5}$$

As the Black-Scholes market is complete, any derivative can be priced using (2.5), the main result is the pricing formula for European calls and puts – the *Black-Scholes pricing formula*.

Let us denote the price of a European call option with expiration date T and strike price K implied by Black-Scholes market assumptions by C_t^{BS}. Then

$$C_t^{BS}(S_t, K, \tau, r, \sigma) = S_t\Phi(d_1) - Ke^{-r\tau}\Phi(d_2) \tag{2.6}$$

where $d_1 = \frac{\ln(S_t/K)+(r+1/2\sigma^2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, and $\tau = T - t$ is *time to maturity*. The original proof presented in Black and Scholes [1973] was motivated by the construction of a replicating portfolio containing the option and the underlying stock S_t, whose return needs to be equal to the risk-free interest rate r. More generally considering a (path-independent) derivative H with payoff-function $\psi(S_T)$ and denoting its price function by $h(t, S_t)$ h is obtained as a solution of the following partial differential equation:

$$\frac{\partial h}{\partial t} + ts\frac{\partial h}{\partial s} + 0.5\sigma^2 s^2 \frac{\partial^2 h}{\partial s^2} - rh = 0 \tag{2.7}$$

with the boundary condition $h(T, s) = \psi(s)$. For a more detailed discussion see Fengler [2005a] or Hafner [2004] among others. The equation (2.7) is also referred to as Black-Scholes partial differential equation.

The corresponding price of a European Put option P_t can be obtained from the *put-call parity* :

$$P_t(S_t, K, \tau, r, \sigma) = C_t(S_t, K, \tau, r, \sigma) - S_t + e^{-\tau r}K. \tag{2.8}$$

For a proof of (2.8) see Musiela and Rutkowski [1997].

Until now only the no-dividend case was considered. In the case that the underlying of a European option yields a continuous dividend δ over the interval $[0, T]$ the Black-Scholes formula can be generalized to

$$C_t^{BS}(S_t, K, \tau, r, \sigma, \delta) = e^{-\delta\tau}S_t\Phi(d_1) - Ke^{-r\tau}\Phi(d_2) \tag{2.9}$$

where $d_1 = \frac{\ln(S_t/K)+(r-\delta+1/2\sigma^2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, and $\tau = T - t$ is *time to maturity* and the Put-Call parity (2.8) becomes:

$$P_t(S_t, K, \tau, r, \sigma) = C_t(S_t, K, \tau, r, \sigma) - e^{-\tau\delta}S_t + e^{-\tau r}K. \tag{2.10}$$

The next concept that is connected to financial market models (in our case Black-Scholes) are the *Greeks*. The sensitivity of a financial instrument (or even a portfolio) with respect to parameters like spot price (S), volatility (σ), interest rate (r) are denoted by different letters of the Greek alphabet like delta, vega, rho. More precisely the sensitivity is measured in terms of derivatives with respect to the parameters, e.g. the Greeks of the call option whose price is denoted by $C_t = C_t^{BS}$ are (in the Black-Scholes model):

- Delta: $\frac{\partial C_t^{BS}}{\partial S} = \Phi(d_1)$
- Vega: $\frac{\partial C_t^{BS}}{\partial \sigma} = \frac{\phi(d_1)}{S_t\sigma\sqrt{\tau}}$
- Rho: $\frac{\partial C_t^{BS}}{\partial r} = e^{-r\tau}\tau K\Phi(d_2)$.

The Greeks for the Put options with price P_t^{BS} are calculated similarly:

- Delta: $\frac{\partial P_t^{BS}}{\partial S} = \Phi(d_1) - 1$
- Vega: $\frac{\partial P_t^{BS}}{\partial \sigma} = \frac{\phi(d_1)}{S_t\sigma\sqrt{\tau}}$
- Rho: $\frac{\partial P_t^{BS}}{\partial r} = e^{-r\tau}\tau K(\Phi(d_2) - 1)$

There are further Greeks that are commonly considered in financial practice, however these will not be discussed in the following text and hence are omitted at this place. The Greeks are very important for hedging – the management of the risk of a portfolio position in practice. The general idea is to make a portfolio robust, i.e. insensitive w.r.t. the parameters, for example a portfolio with zero delta (delta-neutral portfolio) means that the investor is protected against the change of the spot, however, some trading strategy might also be based on the portfolio's sensitivity to the change of some parameter, for example if the investor expects the change of this parameter in near future. For a deeper discussion of practical aspects of Greeks, see Hull [2003].

2.2 Generalizations of the Black-Scholes Market

It is necessary to mention that the Black-Scholes model is nowadays considered to be too simple to capture all of the phenomena observed in real financial data. It will be discussed in the next chapters that for example the assumption of a constant volatility seems to be violated in real markets, the

log-returns seem to have "fatter tails" (the large returns or losses are more probable) than normal. One possibility how to deal with this problems is to generalize the Black-Scholes model.

The most common generalizations will be introduced. All of them can be essentially described by the generalization of the diffusion equation (2.2):

- **Merton Model** – Merton [1976] argues that the price process of a stock might be affected by a sudden shock, e.g. if important information becomes public. This is modeled by introducing jumps into the dynamics of S_t,

  $$dS_t = S_t\mu dt + \sigma S_t dW_t + S_t dZ_t$$

  where Z_t is a compounded Poisson process, where the jump times follow a Poisson process N_t independent of W_t and the log-jumps are i.i.d. normally distributed and are independent of N_t and W_t. This model is also referred to as *Jump diffusion model*.

- **Heston Model** attacks the constant volatility parameter σ in (2.2) by substituting the σ by an unobserved stochastic process σ_t, the stochastic process σ_t may be designed in different ways, the model proposed by Heston [1993] has the following structure:

  $$\begin{aligned} dS_t &= S_t\mu dt + S_t\sqrt{\sigma_t}W_t^{(1)} \\ d\sigma_t &= \kappa(\theta - \sigma_t)dt + \sigma^{(2)}\sqrt{\sigma_t}dW_t^{(2)} \end{aligned}$$

  where $W_t^{(1)}$ and $W_t^{(2)}$ denote (possibly correlated) Wiener processes.

The κ is the mean reverting parameter, θ the mean level of volatility and $\sigma^{(2)}$ stands for the volatility of the volatility process. This model is also refereed to as *Stochastic Volatility model*

Of course, further generalizations are possible, e.g. a combination of the Merton and the Heston model was proposed by Bates [1996].

Although, as argued above, the Black-Scholes model is not used for pricing today it is still commonly used and the reasons for its popularity will be illustrated by the introduction of the concept of *implied volatility* in the section 2.3.

2.3 Black Scholes Implied Volatility

The concept of Implied Volatility (IV), first proposed by Latané and Rendelman [1976] is based on inversion of the formula (2.6).

Assuming the no-dividend case (i.e. $\delta = 0$) for simplicity, the parameters of the Black-Scholes formula (2.9) – current time, underlying stock, interest rate, time-to-maturity are directly observable on the market, the risk-free rate can be approximated by the inter-bank offered rates published by the central banks (e.g. EURIBOR for the EURO-Zone) or by government bonds. Given this set of parameters (2.9) maps the option price to the volatility parameter σ. This mapping is monotone for $\sigma > 0$. Consequently, observing the market price of an option a volatility parameter σ *implied* by the Black-Scholes formula can be determined.

The *implied volatility* $\tilde{\sigma}$ is defined as the volatility σ, which makes the BS price C_t^{BS} equal to the price $\tilde{C}_t$ observed on the market. For a single asset, one obtains at each time point t a two-dimensional IV surface (IVS) – $\tilde{\sigma}_t(K, \tau)$. One can use the same argument and calculate an implied volatility using Put-option prices.

Please note, that due to the non-linearity of (2.9) the implied volatility has to be calculated by some numerical iterative procedure, usually the Newton-Raphson method is used.

It is necessary to note that under the assumption of arbitrage free, frictionless market and no-dividend-case the IV calculated from the Call options or Put options have to yield the same result, see Hafner [2004] p. 35 among others.

2.4 State Price Density

An arbitrage-free and frictionless market with a risk-free saving account is assumed again. Considering a Call option with strike K and maturity date T on an underlying with price process S_t, i.e. a claim H with the pay-off function $\Phi(S_T) = \max\{S_T - K, 0\}$, the pricing formula (2.5) yields for the price $C_t(K, T)$ at the current time-point t:

$$\begin{aligned} C_t(K,T) &= e^{-r(T-t)} E_Q\left(H|\mathcal{F}_t\right) \quad \text{for all } t \in [0,T] \\ &= e^{-r(T-t)} \int_0^\infty \Phi(s) q_{t,S_T}(s) ds \\ &= e^{-r(T-t)} \int_0^\infty \max\{s-K,0\} q_{t,S_T}(s) ds \end{aligned}$$

where q_{t,S_T} denotes the risk neutral density of S_T at the current point t. The concept of *state price density* (SPD) is also called *risk neutral density*

(RND), see formula (2.1) and enclosed discussion. The theorem of Breeden-Litzenberger gives a way of estimating the risk-neutral density as the second derivative of the $C_t(K,T)$ w.r.t. K:

$$q_{t,S_T}(K) = e^{-r(T-t)} \frac{\partial^2 C_t(K,T)}{\partial K^2} \tag{2.11}$$

Note that equation (2.11) is an abbreviated version of

$$q_{t,S_T}(s,\tau) = e^{-r(T-t)} \frac{\partial^2 C_t(K,T)}{\partial K^2}\bigg|_{K=s},$$

however in order to simplify the following formulas the notation in form of (2.11) is used.

The connection of the SPD with the IV can be obtained by connecting (2.11) and (2.6) after some further algebra (see B.1):

$$q_{t,\tau}(K) = e^{r\tau} S_t \phi(d_1) \sqrt{\tau} \left\{ \frac{1}{K^2 \sigma \tau} + \frac{\partial \sigma}{\partial K} \frac{2d_1}{\sqrt{\tau} K \sigma} + \frac{\partial^2 \sigma}{\partial K^2} + \left(\frac{\partial \sigma}{\partial K} \right)^2 \left[\frac{d_1 d_2}{\sigma} \right] \right\} \tag{2.12}$$

or equivalently

$$q_{t,\tau}(K) = \phi(d_2) \left\{ \frac{1}{K \sigma \sqrt{\tau}} + \frac{\partial \sigma}{\partial K} \frac{2d_1}{\sqrt{\tau} \sigma} + \frac{\partial^2 \sigma}{\partial K^2} + \left(\frac{\partial \sigma}{\partial K} \right)^2 \frac{d_1 d_2}{\sigma} \right\}. \tag{2.13}$$

Breeden and Litzenberger [1978] motivates the concept of SPD by " ... deriving the prices of primitive securities (security paying 1$ upon a given state of the word at a given date) from the European call-options ...", the name state-price density can also be motivated by following formula:

$$P(S_T \in [K_l, K_u] | S_t) = \int_{K_l}^{K_u} q_{t,S_T}(s,\tau) ds. \tag{2.14}$$

Hence the SPD derived from option market data is a forecast of the density of future values of the underlying asset and can be directly used e.g. for constructing prediction intervals. Figure 2.4, borrowed from Härdle and Hlávka [2005], shows the 0.95 prediction intervals based on the SPD estimated from option prices, the grey lines are the 0.025 and 0.975 quantile lines of the estimated SPD at the $\tau = 45$ days, the black line is the true DAX level at future date. For technical details of the calculation of these intervals see Härdle and Hlávka [2005].

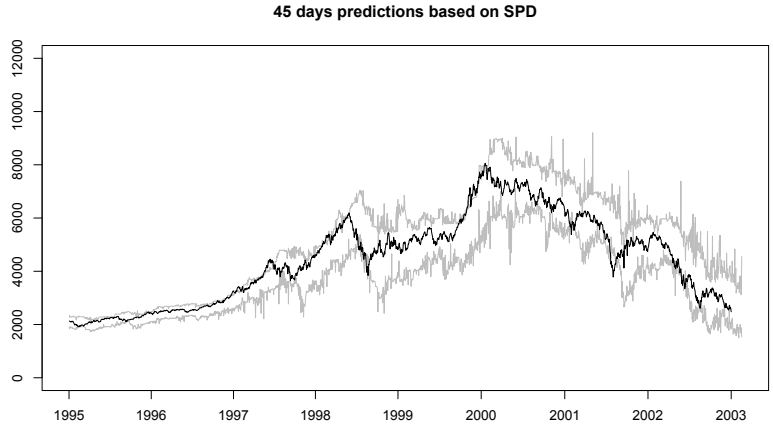

Figure 2.1: Prediction intervals for 45 days future DAX value based on the estimated SPD. Grey lines are the quantile lines (0.025 and 0.975). The black line is the true DAX level on the future date.

For a derivative with a payoff-function $H(S_T)$ depending on the asset-price S_T at the expiration date T, the well known arbitrage-free pricing formula, (2.5) yields the price $\Pi_t(H)$ given by:

$$\Pi_t(H) = e^{-r(T-t)} E_Q\left(H|\mathcal{F}_t\right) \text{ for all } t \in [0,T], \tag{2.15}$$

where Q is the so called equivalent martingale measure. Written in terms of the SPD, this yields

$$\Pi_t(H) = e^{-r(T-t)} \int_0^\infty H(s) q_{t,S_T}(s,\tau) ds \quad \text{for all } t \in [0,T], \tag{2.16}$$

The last formula is of great practical importance, since the SPD q_{t,S_T} can be estimated from the liquid European styled Call and Put options and the estimated SPD can be used for pricing non-liquid exotic derivatives, such as digital options, see Brunner and Hafner [2003] for further examples.

2.4.1 No-arbitrage Conditions Implied by SPD

As argued above the SPD (q_{t,S_T}) should be a probability density function if there is *no-arbitrage*, thus the SPD should be non-negative:

$$q_{t,S_T}(s) = e^{-r(T-t)} \frac{\partial^2 C_t(K,T)}{\partial K^2} \geq 0, \;\; s \in [0,\infty) \tag{2.17}$$

and the SPD should integrate to one over its support:

$$\int_0^\infty q_{t,S_T}(s)ds = \int_0^\infty e^{-r(T-t)}\frac{\partial^2 C_t(K,T)}{\partial K^2}dK = 1. \tag{2.18}$$

In addition to the two conditions, the SPD should reprice the call-options:

$$\int_0^\infty \max\{s-K,0\}q_{t,S_T}(s)ds = e^{r(T-t)}C_t(K,T),\ K \geq 0. \tag{2.19}$$

The condition (2.18) is referred to as martingale condition.

Some further conditions on the Call-price function and/or IV-surface obtained from general arbitrage considerations can be found in Brunner and Hafner [2003] or Fengler [2005a].

2.4.2 Total Variance

Conditions introduced in section 2.4.1 are mainly conditions on the shape of the call-price function and the IVS as functions of K, i.e. conditions essentially obtained by the construction of portfolios using options with the same maturity. In the "τ-direction" (time to maturity) no sufficient condition is known. However it is clear that the price function of an American call option with the same strike must be a non-decreasing function of time-to-maturity (the option with smaller maturity is "included" in the longer). As argued by Merton [1973] the price-function of European Call options should have the same property, essentially a *calendar call spread*, i.e. a portfolio containing two (European) call options with maturity T_1, T_2 should have a nonnegative value, hence $C_t(K,T_2) \geq C_t(e^{-r(T_2-T_1)}K, T_1)$, for a proof see Brunner and Hafner [2003].

Arbitrage-free conditions for the IVS in the τ-direction can be obtained by considering the *total variance* w:

$$w(K,\tau) \stackrel{\text{def}}{=} \sigma^2(K,\tau)\tau. \tag{2.20}$$

Kahalé [2004] considers a zero-dividend and zero-interest rate case and argues that $w(K,\tau)$ should be a strictly increasing function in τ for fixed K. Fengler [2005a] generalizes these arguments into the time-dependent (deterministic) interest rate case and shows that $w(\bullet,\tau)$ should be strictly increasing for fixed κ in the so-called *future moneyness*

$$\kappa \stackrel{\text{def}}{=} K/F_{t,\tau} \tag{2.21}$$

The $F_{t,\tau}$ is the *forward price*. Using the notation introduced for option a forward contract is an agreement negotiated at the day t to trade (buy or sell)

an underlying with price process S_t at a future date $t+\tau$ for a predefined price K. The forward price $F_{t,\tau}$ is the delivery price that makes the contract have zero value at the current time t. No-arbitrage arguments and the assumption of existence of a constant risk-free interest rate yield:

$$F_{t,\tau} = S_t e^{r\tau}. \tag{2.22}$$

Note that forward contracts are not standardized and are traded OTC. Standardized forward-like contract traded on exchanges are called *futures*, besides the fact that the futures are written on a standardized underlying the most important difference is the settlement procedure – futures are settled every day during the whole life of the future contract whereas a forward contract is settled at T. The aim is to decrease the counter-party risk. Consequently, the forward and future prices are in general not equal, see chapter 1 and 2 in Hull [2003]. However since this difference is small for the problems presented in this work we will not distinguish between future and forward prices, and call (2.22) futures moneyness, see also section 3.6.1.

The concept of *total variance* finalizes the short introduction into the financial mathematics with focus on the IVS. Please note at this point that all concepts described in this chapter are linked together: the price of an (European call) option can be transformed into the IV by the Black-Scholes formula (2.6) the SPD function can be calculated from both: price function or IVS by formulas (2.11) and (2.12) respectively. The arbitrage-free considerations yield some theoretical properties of the price functions and the IVS, partially based on the SPD and the total-variance.

2.5 Applications

Very frequently a question arise:

"Why is IV important" ?

In particular BS IV is an implicit parameter in a model that cannot be generally considered appropriate for pricing, consequently it might be questionable why this concept is interesting in practice. However as argued in the following paragraphs, the IV has a practical importance and a practical usage.

Current time, the underlying stock price, the interest rate, the time-to-maturity are directly observable on the market, and given this set of parameters (2.9) maps the option price to the volatility parameter σ, this mapping is monotone for $\sigma > 0$. Hence, the *IV is a transformation of the option price.*

IV as standardization of option prices As it will be discussed in chapter 3, option prices and IVs observed on the market are often subject to noise and this transformation yields a variable that tend to be more smooth, see Shimko [1993] and Rosenberg [2000]. Moreover the SPD can be derived from IV using (2.12) and e.g. Bliss and Panigirtzoglou [2002] argues in favor of the SPD estimation based on IVS, see also Brunner and Hafner [2003] for further bibliographic notes on this topic.

IV as a predictor for realized volatility Implied volatility is often understood as an option based estimator of the future *realized volatility*, realized volatility is essentially a standard deviation based on the (compounded) stock returns over a specific (historical) period of time, see Hafner [2004], p. 33. This approach is very popular, for a bibliographic overview, see Fengler [2005b], section 2.10, – there are empirical as well as theoretical arguments that the IV based estimator typically out-performs an estimator based on purely technical (time series based) analysis of historical returns. On the other hand the IV based estimation is biased (upward), as discussed in Britten-Jones and Neuberger [2000]. A more sophisticated approach attacking this problem by considering the *model-free implied volatility* is studied by Jiang and Tian [2003].

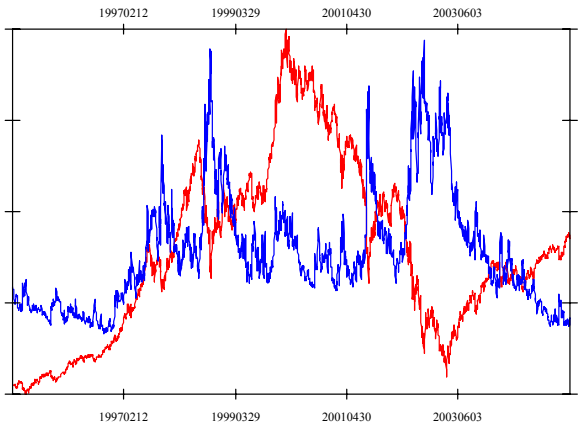

Figure 2.2: Time plot of DAX centered closing level (red dashed line) and centered scaled (multiplied by 10^4) IV (blue solid line) at ATM and maturity 45 days from Jan 1st 1995 to June 30th 2005. IV estimated by local polynomial estimator.

IV as an indicator of market uncertainity In fact, understanding the IV as an indicator of market uncertainity is motivated by the previous paragraph (IV as a predictor of realized volatility): high values of IV correspond to market believes on high future realized volatility and consequently higher fluctuations of the stock price process in the future. Practically even more relevant is the *leverage effect* phenomenon observed in the financial markets – the negative correlation between stock prices returns and their volatilities returns – if the stock price is dropping the IV is raising. This can be used both for hedging and for speculative strategies. For these reasons modern financial markets contain volatility indexes that allow to trade the volatility "directly". In the EUREX exchange the index is named VDAX and VDAX-NEW (trademarks of Deutsche Börse AG), the VDAX is the volatility index and seems to be essentially based on the "at-the-money" (ATM see section 3.5.2 for definition) average of IVS and hence corresponds to the understanding of the BS IV as predictor of future realized volatility, the new volatility index VDAX-NEW seems to be motivated by the model-free IV, for details on construction and differences, see Börse [2006a].

Figure 2.2 illustrates the leverage effect – scaled (multiplied by 10^4) centered ATM IV (blue solid line) at maturity $\tau = 0.125$ (45 days) against the closing level of the DAX index, from Jan 1st 1995 to June 30th 2005. The IV is estimated by a local polynomial estimator, see chapter 3 for details. We can see that the downward drops in the DAX levels corresponds to the upward jumps in the IV. The overall correlation between IV returns and the DAX returns for the whole period (Jan 1995 - June 2005) is -0.51.

IVS, Local Volatility and Pricing Clearly, the non-flat IVS is also an issue in the pricing of financial derivatives, the connection between IVS and path independent options can be setup via the SPD and formula (2.12). Understanding volatility structure and dynamics becomes even more important for path dependent financial derivatives.

An important concept connected to the option pricing is *local volatility*. The concept of local volatility, introduced by Dupire [1994], is defined as expected value of the instantaneous volatility $\sigma^2_{K,T}(S_t, t)$ w.r.t. the risk-neutral measure Q conditional on future state of the asset $S_T = K$, and filtration $\mathcal{F}_t$:

$$\sigma^2_{K,T}(S_t, t) \stackrel{\text{def}}{=} E_Q\{\sigma^2 | S_T = K, \mathcal{F}_t\},$$

where σ^2 is the instantanuous volatility as used in the difution equation (2.2).

Basic difference in the interpretation of the *local* volatility and the *implied* volatility is that the *local* is interpreted as the expectation of the market participants about the future instantaneous volatility whereas the implied

can be interpreted as the expected average volatility between t and T. From these reasons the local volatility is also referred to as *forward* volatility.

Similarly to the IV the local volatility can be obtained from the call prices via (Dupire formula) Dupire [1994]:

$$\sigma^2_{K,T}(S_t,t) = 2\frac{\frac{\partial C_t(K,T)}{\partial T} + \delta C_t(K,T) + (r-\delta)K\frac{\partial C_t(K,T)}{\partial K}}{K^2\frac{\partial^2 C_t(K,T)}{\partial K^2}} \tag{2.23}$$

Since the IV is a transformation of an option price, Dupire formula (2.23) can be rewritten in terms of the IV, see Fengler [2005b], section 3.5. It should be noted that under no-arbitrage and using some further results in the financial theory is $\sigma^2_{K,T}(S_t,t)$ positive. However call-price function and its derivatives need to be estimated in practice from the data observed on the market, see chapter 3 where the no-arbitrage condition does not need to be necessarily fulfilled, e.g. may lead to negative SPD and consequently negative denominator in (2.23) and the no-arbitrage issues needs to be considered if the estimated call-price function or IVs is used to calculate the local volatility through the Dupire formula (2.23), see also section 3.8 for comments on the connected issues.

Chapter 3

Nonparametric Regression and Empirical Finance

The implied volatility, SPD and option-price-function, introduced in the chapter 2 can be all considered as functions of time-to-maturity and strike price. Clearly, from the theoretical point of view, an option can be constructed for arbitrary time-to-maturity or strike price. However, only options on some discrete grid of these parameters can be observed on the market for a given time. This is caused by the market mechanism and conventions and by different trading intensity for different parameters (liquidity issues). For similar reasons observed data can be contaminated by some (observational) noise. This chapter focuses on the problem of estimating the "true" function (e.g. option-price or implied volatility) from the noisy data observed on the discrete grid.

The chapter is organized as follows: section 3.1 discusses the theoretical aspects of nonparametric regression techniques, focuses on local polynomial estimators, discussed in section 3.2 and on the functional basis expansion techniques in section 3.3. Kernel density estimates are introduced in section 3.4.

After introducing nonparametric techniques and discussing basic theoretical properties, the issues of estimating IV-function and surfaces, calculating the IVs from the raw data (option prices, interest rates and recovering of spot prices) and liquidity are discussed in section 3.5.

Observation model assumed by estimating the IVS is introduced in section 3.6, typically observed phenomena of the statically estimated IVS are summarized in section 3.7. Estimating the IV-function and IVS by combining the nonparametric techniques with the results obtained by arbitrage-free theory is introduced in the section 3.8.

Concluding remarks and outlook can be found in section 3.9

3.1 Nonparametric Regression

Consider a random sample (Y_i, t_i), $i = 1, \ldots, T$. A regression model for the purposes of this chapter is defined as a set of functions $(X(\bullet), \sigma(\bullet))$, where $X(\bullet)$ is a real valued function, $\sigma(\bullet)$ is a positive real valued function and the following holds:

$$Y_i = X(t_i) + \varepsilon_i \text{ for } i = 1 \ldots T \tag{3.1}$$

where $\varepsilon_i, i = 1 \ldots T$ are zero mean i.i.d. random variables and $Var(\varepsilon_i) = \sigma(t_i), i = 1 \ldots T$, $\varepsilon_1, \ldots, \varepsilon_T$, $t_1, \ldots, t_T$ are independent.

If not written others, in the following text, the support of X and σ is $[0, 1]$ (this can be assumed without loss of generality). Please note that small inconsistency in the notation occurred – T used for expiration-time of an financial derivatives is used here for number of observation in the model (3.1). Since, it should be clear in which role the T is employed in the following text we will, in order to be consistent with the notation with further chapters, allow ourself this inaccuracy.

The variables t_i (possibly multi-dimensional) are called *explanatory variables*, also called *design points*. The whole vector $\mathbf{t} = (t_1, \ldots, t_T)$ is referred to as *design* or *design matrix* for multi-dimensional case.

Please note that the design $(t_1, \ldots, t_T)$ where t_i are random variables is referred to as *random design*, in this case the marginal density of t_i is denoted by f_t (*design density*). In case where the t_i are deterministic variables (can be chosen or are known in advance) the design is called *fixed*.

Random variables $\mathbf{Y} = (Y_1, \ldots, Y_T)$ are called *dependent variables*. The function X is the *regression function*, often called *mean function*, the function σ is the *variance function*.

The main aim of this chapter is to estimate the mean function X in the (3.1). Two basic types of estimating the regression function are possible, first - the *parametric approach* where the X is known up to a finite set of parameters, e.g. well known linear regression where X is assumed to be a linear function of the dependent variables and only the corresponding coefficient and intercept need to be estimated. A parametric model can be estimated typically fast, by low computational costs (for a textbook on parametric modeling see HUMAK [1997] among others), however in practice the assumption of "knowing the structure" of the regression curve is often not justified. There is an alternative class of methods – *nonparametric modeling* that relaxes the assumption on the regression curve and overcomes this specification problem, however by paying tribute on the precision of the estimate and computational speed.

There are many nonparametric methods in regression, proposed by dif-

ferent authors, for overview see Härdle [1990] or Ruppert et al. [2003] among others.

In the next two sections, two types of nonparametric estimation techniques are presented, local polynomial estimation in section 3.2 and functional basis approach in section 3.3.

3.2 Local Polynomials

The local polynomial estimation can be motivated by well known Taylor expansion.

Assume that X has a $(p+1)$th continuous derivative, then the regression function X at the point t_i can by approximated by

$$X(t) + X^{(1)}(t)(t - t_i) + \cdots + X^{(p)}(t)(t - t_i)^p \frac{1}{p!} \tag{3.2}$$

with the approximation error of order $(t-t_i)^{p+1}$, $X^{(j)}$ stands for jth derivative of X.

The derivatives of regression function, more precisely the factors $\frac{X^{(l)}(t)}{l!}$, $l = 0, \ldots, p$ in (3.2), can be understood as parameters (depending on the point t) and can be estimated, e.g. by *weighted least squares*-(WLS), see below.

This approximation is, however, appropriate only in the neighborhood of the point t. The localization of the approximation (3.2) is achieved by weighting function $\mathcal{K}$, typically defined on the support $[-1, 1]$ and re-scaled by parameter h – "width of the neighborhood". Hence the estimate of regression function X at the point t is given by minimization of the following criterion:

$$\hat{\beta}(t, h, p, \mathcal{K}) = argmin_\beta \sum_{i=1}^{T} [Y_i - \beta_0 - \beta_1(t_i - t) - \cdots - \beta_p(t_i - t)^p]^2 \mathcal{K}\left(\frac{t_i - t}{h}\right). \tag{3.3}$$

The weighting function $\mathcal{K}(\bullet)$ is often referred to as *Kernel function*, typically a probability density function is used. The scaling factor h is called bandwidth.

Comparison of (3.2) and (3.3) yields $\beta_0 = \hat{X}(t)$, where $\hat{X}(t)$ denotes the estimate of the regression function at the point t, moreover

$$\hat{X}^{(\nu)}(t) = \nu! \hat{\beta}_\nu(t, h, p, \mathcal{K}), \tag{3.4}$$

where $\hat{\beta}_\nu$ denotes an estimate of the νth derivative of the regression function. Hence the local polynomial estimate delivers not only the estimate of the regression function but also its derivatives up to the order p.

Note that the local polynomial estimate can be obtained by WLS method only if the neighborhood of point t – $[t-h, t+h]$ contains at least $p+1$ design points t_i with $Y_i \neq Y_j$.

An interesting special case is the case $p = 0$, i.e. local constant estimate. In this case an estimate of $X(t)$ can be written in the following form :

$$\hat{X}_h^{NW}(t) = \frac{\sum_{i=1}^{T} Y_i \mathcal{K}(\frac{t-t_i}{h})}{\sum_{i=1}^{T} \mathcal{K}(\frac{t-t_i}{h})}. \tag{3.5}$$

The estimate (3.5) is often called Nadaraya-Watson (kernel) estimator (NW).

A detailed discussion of the theoretical properties of the local polynomial estimates is presented in Fan and Gijbels [1996], these results imply that under some further regularity condition the leading term of the asymptotic conditional variance of $\hat{X}^{(\nu)}(t)$ is of order $\mathcal{O}_p(T^{-1}h^{-(1+2\nu)})$ and is proportional to the true variance function $\sigma^2(t)$ and $(1/f_t(t))$.

The leading term of asymptotic conditional bias is of order $\mathcal{O}_p(h^{p+1-\nu})$ and is proportional

- to $X^{(p+1)}(t)$ and $\frac{1}{f_t(t)}$ for $p-\nu$ odd

- $\mathcal{O}_p(h^{p+2-\nu})$ and proportional to $X^{(p+2)}(t)$, $X^{(p+1)}(t)\frac{f_t'(t)}{f_t(t)}$ for $p-\nu$ even,

furthermore both, variance and bias, depend on the kernel function $\mathcal{K}$. For details see Theorem 3.1 in Fan and Gijbels [1996].

The formulas for asymptotic conditional bias and variance give insight into the role of the bandwidth h. A large h increase the bias and decrease the variance, small h vice versa. In general, the choice of the bandwidth – the *bandwidth-selection* methods are based on the balancing of the bias and variance. The bandwidth plays a crucial role in the practical usage of the local polynomials and a variety of automated selection rules have been proposed.

There are two basic approaches: first is the *global bandwidth* choice – a bandwidth h valid for all points $t \in [0, 1]$, essentially chooses the bandwidth by minimizing the estimate of the MISE of the estimate e.g. *cross validation* argument, *plug-in methods*, or by introducing the *penalizing functions*, for an overview see Härdle et al. [2004], section 4.3. A more sophisticated approach – is the *local bandwidth choice* – chosen essentially by minimizing the MSE individually for each t where the estimate is constructed, see Härdle [1990] or adaptive methods recently proposed by Spokoiny [2006].

Unfortunately it seems that there is no "perfect bandwidth selection" method that always out-performs all others and often even a setup of the bandwidth chosen by experimenters visual choice seems to have a similar quality by estimating the "optimal" bandwidth.

There are two other parameters of the local polynomial estimate – the kernel function $\mathcal{K}$ and the order of the polynomial p. The choice of kernel function has not a crucial role in the practice, see Härdle et al. [2004], section 3.4 among others, typically a probability density function is used, popular choices are:

- *Uniform kernel* $\mathcal{K}(u) = \frac{1}{2}\mathbf{1}(|u| \leq 1)$,
- *Epanechnikov kernel* $\mathcal{K}(u) = \frac{3}{4}(1-u^2)\mathbf{1}(|u| \leq 1)$,
- *Quartic kernel* $\mathcal{K}(u) = \frac{15}{16}(1-u^2)^2\mathbf{1}(|u| \leq 1)$,
- *Gaussian kernel* $\mathcal{K}(u) = (2\pi)^{-1/2}\exp\left(\frac{-u^2}{2}\right)$.

The order of the polynomial however is again of particular importance in practice, clearly p has to fulfill $p \geq \nu$ and, as already observed by introducing the leading terms of errors, two cases $p-\nu$ odd and even are distinguished, generally $p-\nu$ odd out-performs the estimate with $p-\nu$ even. Another issue on this topic is the computational complications connected to the usage of higher order polynomials, optimization criterion (3.3) is in practice solved by

$$\hat{\beta} = argmin_{\beta}(\mathbf{Y} - \mathbf{t}_{\mathbf{t},\mathbf{T},\mathbf{p}}\beta)^{\top}\mathbf{W}_{\mathbf{t},\mathbf{T},\mathbf{p},\mathbf{h},\mathcal{K}}(\mathbf{Y} - \mathbf{t}_{\mathbf{t},\mathbf{T},\mathbf{p}}\beta) \tag{3.6}$$

where

$$\mathbf{t}_{\mathbf{t},\mathbf{T},\mathbf{p}} = \begin{pmatrix} 1 & t_1 - t & (t_1-t)^2 & \dots & (t_1-t)^p \\ 1 & t_2 - t & (t_2-t)^2 & \dots & (t_2-t)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_T - t & (t_T-t)^2 & \dots & (t_T-t)^p \end{pmatrix} \tag{3.7}$$

$$\mathbf{W}_{\mathbf{t},\mathbf{T},\mathbf{p},\mathbf{h},\mathcal{K}} = \begin{pmatrix} \mathcal{K}(\frac{t_1-t}{h}) & 0 & \dots & 0 \\ 0 & \mathcal{K}(\frac{t_2-t}{h}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathcal{K}(\frac{t_T-t}{h}) \end{pmatrix}. \tag{3.8}$$

Clearly from (3.7) it is visible that higher order polynomials yield higher computational costs and typically the choice $p = \nu + 1$ is preferred, for detailed discussion see Fan and Gijbels [1996], section 3.3.

Local polynomial estimators belong to an interesting class of smoothing methods – *linear smoothers*, smoothing methods that can be written as locally weighted average of depending variables:

$$\hat{X}(t) = \sum_{i=1}^{T} w_i(t) Y_i. \tag{3.9}$$

For NW-smoothers (3.5) the weights are obviously given by:

$$w_i(t) = \frac{\mathcal{K}(\frac{t-t_i}{h})}{\sum_{i=1}^{T} \mathcal{K}(\frac{t-t_i}{h})}.$$

For local polynomial estimator estimator $\hat{\beta}_\nu$ can be written in the form:

$$\hat{\beta}_\nu = \sum_{i=1}^{T} w_\nu \left(\frac{t_i - t}{h} \right) Y_i$$

$w_\nu^T(u) = \mathbf{e}_{\nu+1}^\top ({\mathbf{t}_{\mathbf{t},\mathbf{n},\mathbf{p}}}^\top \mathbf{W}_{\mathbf{t},\mathbf{T},\mathbf{p},\mathbf{h},\mathcal{K}} \mathbf{t}_{\mathbf{t},\mathbf{n},\mathbf{p}})^{-1} \{1, uh, \ldots, (uh)^p\}^\top \mathcal{K}(u)/h$, for $\nu = 0, \ldots, p$.

Note that, a local polynomial estimator can be rewritten into the form of a kernel estimator with so-called equivalent (higher order) kernel with similar asymptotic properties. The "higher order kernel representations" are often used by the investigations of the theoretical properties of the estimates based on the local polynomials. However following the arguments summarized in Fan and Gijbels [1996], section 3.2.2, local polynomial estimates can better adapt to various designs than kernel methods based on the higher order kernels and have better properties on the boundaries (so-called "automatic boundary carpentry"-property of the local polynomials).

The idea of the local polynomial estimation can be extended to multi dimensional regression problems in straight forward way, application and some comments on the two-dimensional smoothing (surface smoothing) will be given in section 3.8.2, see also Fan and Gijbels [1996], chapter 7. However, there are two basic problems connected to the nonparametric approach in the higher dimensions. First, there is a practical problem by visualization of the results – in general case an output of the local smoother is a set of functional values on a dense grid of dimension equal to the dimension of the regressor (t). Typical visualization for one dimensional regressor is simple graph of the function, often together with its derivatives. For two dimensional regressor, an output of the local polynomial estimation are functional values on a two-dimensional grid, a wide range of visualization techniques for two

dimensional data has been proposed here, starting from surface plots, contour plots, and many others, see Härdle et al. [2000] section 3.3. For visualization of the three dimensional functions (regression problem with three dimensional regressor) some proposal has been made e.g. three dimensional contour plots, for dimensions above three, the user is essentially restricted to visualize only a lower-dimensional projection of the results. Even more serious is the so called *curse-of-dimensions*, that refers to exponentially decay of precision (in terms of MISE) by increasing dimension of the regressor. The curse-of-dimensions can be illustrated by the problem of filling a d dimensional unit cube with an equidistant grid of length n – for one dimension only n points are needed, in two dimension n^2, in general case, for d-dimensional cube n^d. Only additional information on the structure of regression problem can avoid the curse-of-dimension, this leads among others to so called semiparametric regression models, for overview see Härdle et al. [2004] and Ruppert et al. [2003] among others.

Application of the nonparametric techniques to the variance estimation (estimating the σ in 3.1) as well as nonparametric specification tests – lack-of-fit tests are deeply discussed in Hart [1997].

3.3 Functional Basis Expansion

A popular approach in regression modeling is the functional basis expansion technique. This approach can be motivated by well known facts that any function continuous on an interval J can be approximated arbitrary well by the polynomials or that any periodic function on an interval J can be expanded by Fourier series.

Consider a functional basis on an interval J, denote it by $\{\theta_1, \theta_2, \dots, \}$ and assume that the function X belongs to the linear space spanned by θ_l, $l = 1, 2, \dots L$:

$$X(t) = \sum_{l=1}^{L} c_l \theta_l(t) = \mathbf{c}^\top \boldsymbol{\theta}(t), \tag{3.10}$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_L)^\top$ and $\mathbf{c} = (c_1, \dots, c_L)^\top$. Inserting (3.10) into the general model (3.1) yields

$$Y_i = \sum_{l=1}^{L} c_{il} \theta_l(t_i) + \varepsilon_i \text{ for } i = 1 \dots T \tag{3.11}$$

There are three prominent examples of functional bases: Fourier, Polynomial
and B-Spline basis.

3.3.1 Fourier Basis

A well known basis for periodic functions on the interval J is the Fourier basis, defined on J by

$$\phi_0(t) = \frac{1}{\sqrt{|J|}}, \tag{3.12}$$

$$\phi_{2r-1}(t) = \frac{1}{\sqrt{|J|/2}} \sin(r\omega t), \tag{3.13}$$

$$\phi_{2r}(t) = \frac{1}{\sqrt{|J|/2}} \cos(r\omega t), \tag{3.14}$$

for $r = 1, \ldots, L/2$, where L is an even integer. The frequency ω determines the period and the length of the interval $|J| = 2\pi/\omega$ (here $J = [0, 1]$ is assumed hence $|J| = 1$).

The Fourier basis defined above is an orthonormal basis. The popularity of this basis is based partially on the possibility of fast coefficient calculation by the Fast Fourier Transformation (FFT) Algorithm.

Another important feature of the Fourier series is the existence of continuous derivatives:

$$\begin{aligned} D^m\phi_0(t) &= 0, m \geq 1 \\ D^m\phi_{2r-1}(t) &= (r\omega)^m \sin\left(r\omega t + \frac{m\pi}{2\omega}\right), \ m \geq 0, \\ D^m\phi_{2r}(t) &= (r\omega)^m \cos\left(r\omega t + \frac{m\pi}{2\omega}\right), \ m \geq 0 \end{aligned}$$

for $r = 1, \ldots, L/2$. As it will be obvious from later discussion the possibility of easy calculation of the derivative simplifies the implementation of the estimators based on Fourier series, on the other hand in practice, it is very unrealistic to assume existence of all continuous derivatives, as it is done by Fourier basis for fixed L. Consequently, estimates based on the Fourier series are not performing well for functions with strong local features, like discontinuity points in lower order derivatives.

3.3.2 Polynomial Basis

The polynomial basis, appropriate for non-periodic functions is defined by

$$\theta_l(t) = (t - \omega)^k, k = 0, 1, \ldots, L - 1 \tag{3.15}$$

where ω is a shift parameter. The polynomial functions are easy to calculate, for example by a simple recursion. The calculation of derivatives is also very

simple and fast. However, high order polynomials become too fluctuating especially in the boundaries of J.

Polynomial basis as defined in (3.15) is not orthogonal. However several modified types of polynomial systems, are orthogonal, e.g. Legendre polynomials.

3.3.3 B-Spline Basis

A very popular functional basis for non-periodic data is the B-Spline basis. This basis is defined by a sequence of knots on the interval J and is roughly speaking a basis for piecewise polynomial functions of order K smoothly connected in the knots. More formally, the basis functions are

$$\theta_l(t) = B_{l,K}(t), l = 1, \ldots, m + k - 2 \tag{3.16}$$

where $B_{l,K}$ is l-th B-Spline of order K, for the non-decreasing sequence of knots $\{\tau_i\}_{i=1}^m$ defined by the following recursion scheme:

$$B_{i,1}(t) = \begin{cases} 1, & \text{for } t \in [\tau_i, \tau_{i+1}] \\ 0, & \text{otherwise} \end{cases}$$

$$B_{i,k}(t) = \frac{t - \tau_i}{\tau_{i+k-1} - \tau_i} B_{i,k-1}(t) + \frac{\tau_{i+k} - t}{\tau_{i+k} - \tau_{i+1}} B_{i+1,k-1}(t)$$

for $i = 1, \ldots, m + k$, $k = 0, \ldots, K$. The number of the basis functions will uniquely be defined by the B-spline order and the number of knots. The advantage of the B-spline basis is its flexibility, relatively fast evaluation of the basis functions and their derivatives.

The detailed discussion of the implementation of the B-spline basis expansion in the statistical software (statistical computing environment XploRe) can be found in Ulbricht [2004].

3.3.4 Approximation and Coefficient Estimation

By fixed number of basis function L, the regression model (3.11) is linear w.r.t. transformed variables $\theta_l(t)$ for $l = 1, \ldots, L$. Hence, a natural candidate for estimating the coefficient vector $\mathbf{c}$ is found by minimizing some loss function. Quadratic loss function is most commonly used and this choice leads to the generalized least squares (GLS) estimator:

$$\hat{\mathbf{c}} = \left\{ \boldsymbol{\theta}(\mathbf{t})^\top \Sigma^{-1} \boldsymbol{\theta}(\mathbf{t}) \right\}^{-1} \theta(\mathbf{t})^\top \Sigma^{-1} \mathbf{Y} \tag{3.17}$$

where Σ is the covariance matrix of the residuals and $\boldsymbol{\theta}(\mathbf{t})$ is the matrix with elements $\theta_l(t_i)$ for $l = 1, \ldots, L$ and $i = 1, \ldots, T$.

In the case of heteroscedastic independent errors considered here, Σ is diagonal matrix with $\sigma(t_i)$ on the diagonal. Since Σ is typically unknown, an estimate $\hat{\Sigma}$ has to be used, for details see Judge et al. [1988].

In the nonparametric regression, L is however unkown and needs to be estimated from the data – important question for practitioners is how many functions should be used in the basis expansion. This problem is essentially equivalent to the bandwidth-choice in the local polynomials discussed in section 3.2. Although, as stated by Ramsay and Silverman [2002], even a subjective selection of the smoothing parameter leads usually to a reasonable choice, from a statistical point of view an automated (data driven) selection is needed. In the simplest case of e.g. Fourier basis without using additional regularization we need to set just the L. This can be done easily using Cross-Validation, Generalized Cross Validation or other similar criteria described in Härdle [1990] among others. More complicated is the parameter-choice in the case of B-splines basis – in practice we need to choose the knot sequence in addition to the number of functions. In some special applications the choice of knot points is naturally given by the underlying problem. One practical rule of thumb can be a good starting point: set at least 3 knots in the neighborhood of the "interesting" point of the function, e.g. around the expected extreme-point or another change in the function. For an bibliographic overview for automated knot selection for B-Splines, see Ruppert et al. [2003], section 3.4.

Penalized Spline Regression

Another approach is the *penalized spline regression*. The basic idea is to combine LS squares (or GLS) criterion with some penalization of the roughness of the resulting functions $\hat{X} = \hat{\mathbf{c}}^\top \boldsymbol{\theta}$.

Defining the roughness penalty as a norm of a function after applying an operator on a (Hilbert) space $\mathcal{H}$, $R(X) \stackrel{\text{def}}{=} \| \mathcal{L}(X) \|^2$, $\mathcal{L} \in \mathcal{H}^*$ we will minimize:

$$\sum_{i=1}^{T} \left(Y_i - \mathbf{c}^\top \boldsymbol{\theta}(t_i)\right)^2 + \alpha \| \mathcal{L}(\mathbf{c}^\top \boldsymbol{\theta}) \|^2 \tag{3.18}$$

where α is a parameter controling the degree of penalization. Clearly $\alpha = 0$ yields the least square regression (3.17). A popular example of the roughness penalty is $\mathcal{L} = \mathcal{D}^2$ where we penalize nonlinearity of the estimated function $\mathbf{c}^\top \boldsymbol{\theta}$. A more general appoach assumes that $\mathcal{L}$ is a linear differential operator, i.e.

$$\mathcal{L} = a_1 \mathcal{D}^1 + a_2 \mathcal{D}^2 + \ldots + a_P \mathcal{D}^P. \tag{3.19}$$

The proper choice of the operator should have background in some additional information about the underlying function. Assume for example that $X \in \mathcal{V}$,

$\mathcal{V} \subset \mathcal{H}$, then we should try to find an operator $\mathcal{L}$ so that $\mathrm{Ker}(\mathcal{L}) = \mathcal{V}$. Doing so we will penalize the coefficients that yield function $\hat{X} = \mathbf{c}^\top \boldsymbol{\theta} \notin \mathcal{V}$.

Clearly, we can write $\mathcal{L}(\mathbf{c}^\top \boldsymbol{\theta}) = \mathbf{c}^\top \mathcal{L}(\boldsymbol{\theta})$, hence for implementation we need to be able just to calculate the function $\mathcal{L}(\theta_l)$.

An alternative approach may be applied in the case where we have additional information about the function of interest transformed into the roughness penalty $\| \mathcal{L} \|$.

The algorithm is as follows:

1. Use a "full" model for the data set, i.e. use $L \approx$ number of observations. Using this basis directly would lead to a highly volatile estimator with a small (zero) bias.

2. Transform additional information about the function into the kernel of some appropriate linear differential operator.

3. Use the roughness penalty approach and estimate "smoothed" coefficients vector $\mathbf{c}$.

For the cubic B-splines basis, the first step corresponds essentially to setting the knots into each design point. If we set $\mathcal{L} = \mathcal{D}^2$, we in fact penalize nonlinear functions, and obtain a special case of the very popular nonparametric technique – smoothing splines. In the third step of the algorithm we can easily set the smoothing parameter by Cross-Validation (CV) or Generalized Cross-Validation (GCV), for details see Hastie et al. [2002]. This method is fully data driven for a given operator $\mathcal{L}$.

Another approach of the coefficient estimation is based on the direct estimation of Fourier coefficients w.r.t. to basis $\boldsymbol{\theta}$.

First assume that the function X is observed directly (without any observation error and for all $t \in \mathrm{supp}(X)$) and $\{\theta_1, \ldots, \theta_L\}$ is a set of orthonormal functions (last can be assumed without loss of generality, since any set of functions can be orthonormalized, e.g. by Gramm-Schmidt orthogonalization method). Then the coefficients c_l can be directly calculated as Fourier coefficients of X w.r.t. θ:

$$c_l = \int X(t)\theta_l(t)dt \tag{3.20}$$

Of course, due to technical reasons, the function X is observed only on some finite grid $\{t_1, t_2, \ldots, t_T\} \in J$:

$$\mathcal{X} \stackrel{\mathrm{def}}{=} \{X(t_1), X(t_2), \ldots, X(t_T)\},$$

where T denotes the numbers of grid points. The coefficients c_l need to be approximated for example by a (linear or spline) interpolation of X denoted by $X_{int}(t)$, plugging $X_{int}(t)$ into (3.20) yields:

$$\hat{c}_l = \int X_{int}(t)\theta_l(t)dt. \tag{3.21}$$

Even more involved is the case, where model (3.1) is assumed. In this case let A_i be a set of disjoint intervals $\sum_{i=1}^{T} A_i = [0,1]$, and $t_i \in A_i$ for $i = 1, \ldots, T$ e.g. $A_1 = [0, 0.5.(t_2 - t_1)]$, $A_i = [0.5.(t_i - t_{i-1}), 0.5.(t_{i+1} - t_i)]$ for $i = 2, \ldots, T-1$ and $A_T = [0.5.(t_T - t_{T-1}), 1]$. Then

$$\begin{aligned} c_l &= \sum_{i=1}^{T} \int_{A_i} X(t)\theta_l(t)dt \\ &\approx \sum_{i=1}^{T} X(t_i) \int_{A_i} \theta_l(t)dt. \end{aligned}$$

The last formula motivates the *orthogonal series estimator*:

$$\hat{c}_l = \sum_{i=1}^{T} Y_i \int_{A_i} \theta_l(t)dt. \tag{3.22}$$

and finally the estimator for X:

$$\hat{X} = \hat{\mathbf{c}}^\top \boldsymbol{\theta} \tag{3.23}$$

The estimator with coefficient calculated by (3.22) is a linear smoother (3.9) with weights

$$w_i(x) = \sum_{l=1}^{L} \int_{A_i} \theta_l(u)du.$$

In particular Härdle [1990] gives an overview on the results on orthogonal series estimations together with proof of consistency for several types of orthogonal systems, Härdle [1984] gives convergence rates for Legendre polynomials, Hart [1997] gives the detailed discussion of the connection between Fourier and kernel type of estimation.

3.4 Density Estimation

In addition to nonparametric regression techniques, selected nonparametric approaches to the density estimation problem are introduced in this section,

the application of these methods is later used in the analysis of the options on DAX (ODAX) and implied volatilities in section 3.6.1. For extended discussion on (nonparametric) density estimation see monographs Silverman [1990] and Scott [1992].

In order to keep the notation consistent, let the t be a random variable with density function f. The problem of estimating the (unknown) f from an i.i.d. sample $(t_1, t_2 \ldots, t_T)$ is known as *density estimation problem*.

Again a wide range of estimation approaches has been proposed. Two streams can be distinguished – estimation under assumption that f is known up to a finite dimensional parameter θ – *parametric approach*, and the approach discussed here – the *nonparametric approach*. It can be seen as a complement of the parametric – we merely assume a certain level of smoothness of f.

Estimation techniques presented in this chapter can be motivated by the fact that, for an interval J, hollowing holds: $P(t \in J) = \int_J f(s)ds$. In fact the well known *histogram* estimates the f essentially by replacing the $P(t \in [a, b))$ by its empirical counterparts on a (equidistant) grid of the data range – the histogram estimate $\hat{f}_h(t)$ for $t \in (a, b])$ is defined as $\hat{f}_h(t) = \frac{1}{Th} \sum_{i=1}^{T} \frac{1}{2} \boldsymbol{I}(t_i \in [a, b))$ where a, b are two subsequent points of a grid and $h = 0.5(b - a)$. Moving from the piece-wise estimation of the density to the local approach (estimating the density $\hat{f}(t)$ *at* the point t) and using a weighting (kernel) function K defined in (3.40) we obtain the kernel density estimate (KDE):

$$\hat{f}_h(t) = \frac{1}{Th} \sum_{i=1}^{T} \mathcal{K}\left(\frac{t - t_i}{h}\right). \tag{3.24}$$

The bandwidth h rules the amount of smoothness of the estimate $\hat{f}_h(t)$. Using straight forward Taylor expansion arguments, and under some further regularity conditions, it can be easily shown that the leading terms of the asymptotic bias of $\hat{f}_h(t)$ is of order $\mathcal{O}(h^2)$ and is proportional to $f''(t)$, the leading term of asymptotic variance is of order $\frac{1}{Th}$ and is proportional to $f(t)$. Thus, h rules the bias-variance tradeoff, similarly to the local polynomial regression. Moreover $\hat{f}_h(t)$ is a consistent estimate of $f(t)$ as long as $h \to 0$, $Th \to \infty$ for $T \to \infty$. The choice of the bandwidth can be done, in general, by estimating the AMSE or AMISE, among others by the *cross-validation* argument, or by estimating the unknown elements (f and f'' or $||f''||$) in AMSE or AMISE respectively, this yield so called *plug-in* bandwidth rules, see Härdle et al. [2004] among others. Note that $\hat{f}_h(t)$ depends on the choice of the kernel function, (however by "standard" choice of the kernel functions

the differences in the MISE are minimal, see Silverman [1990], p.43) and the choice of the particular kernel function in practice can be driven purely by considering computational and differentiability-issues.

Considering the bi-variate case, the density estimation problem is again straight forward, the task is to estimate a density $f(s,t)$ of a 2-dimensional random vector from the sample data t_{ij}, $i = 1, \ldots, T$ and $j = 1, 2$. Formula (3.24) generalizes to:

$$\hat{f}_H(t) = \frac{1}{T \det H} \sum_{i=1}^{T} \mathcal{K}_2 \left(H^{-1} \left[(t,s) - (t_i, s_i) \right] \right), \tag{3.25}$$

where H is a bandwidth matrix and $\mathcal{K}_2$ a two dimensional kernel function. A two-dimensional kernel function can be chosen to be a product of two one-dimensional kernel functions – Product kernels – ($\mathcal{K}_2(t,s) = \mathcal{K}(t)\mathcal{K}(s)$) or by applying the univariate kernel function to the (Euclidean) norm of vector (s,t) – radial symmetric kernels.

The next step to the general d-dimensional can be done using same arguments. Note, that the AMISE (under usual regularity conditions and optimized w.r.t. bandwidths) is of order $T^{-4/(4+d)}$, indicating again the curse of dimensionality.

There is also a connection of the KDE (3.25) and NW regression estimate (3.5). Considering univariate regression model (3.1) we obtain:

$$X(t) = E(Y|t) = \int y f(y|t) dy = \int y \frac{f(y,t)}{f(t)} dy = \frac{\int y f(y,t) dy}{f(t)} \tag{3.26}$$

where $f(y|t)$ is a conditional density of Y given t, $f(y,t)$ the corresponding joint density, $f_t(t)$ the marginal (design) density of t. Estimating the joint density $f(y,t)$ by a two dimensional KDE $\hat{f}_H(y,t)$ estimate (3.25) with product kernel and diagonal bandwidth matrix $H = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}$, we obtain:

$$\hat{f}_{h_1,h_2}(t,y) = \frac{1}{T} \sum_{i=1}^{T} \frac{1}{h_1} \mathcal{K} \left(\frac{t - t_i}{h_1} \right) \frac{1}{h_2} \mathcal{K} \left(\frac{y - Y_i}{h_2} \right). \tag{3.27}$$

Hence after some simple algebra and by considering $\mathcal{K}$ as p.d.f. symetric around zero (this holds for all Kernel functions considered in this thesis):

$$\int y \hat{f}_h(y,t) dy = (hT)^{-1} \sum_{i=1}^{T} Y_i \mathcal{K} \left(\frac{t - t_i}{h} \right). \tag{3.28}$$

Plugging (3.28) together with standard univariate KDE (3.24) for the $f_t(t)$ in (3.26) yields

$$\hat{X}_h^{NW}(t) = \frac{(hT)^{-1} \sum_{i=1}^{T} Y_i \mathcal{K}(\frac{t-t_i}{h})}{(hT)^{-1} \sum_{i=1}^{T} \mathcal{K}(\frac{t-t_i}{h})}, \tag{3.29}$$

and finally (3.5).

The orthogonal series estimators can be also used for estimating the probability density. Since these estimates will not be used in further analysis, we only refer to Silverman [1990], section 2.7, among others.

3.5 IV calculation - Description of Data

As already announced in section 2.3, the implied volatility is derived by inverting the BS pricing formula (2.6). In this section, we discuss the technical details of this process. The first part describes the raw data – interest rates, option prices and prices of the underlying. Approximation of the interest rates (r) by EURIBOR is discussed in section 3.5.1. The most important input data – the option prices (C_t, P_t) are discussed in section 3.5.2, the extraction of the spot price (S_t) in section 3.5.3, where the correction of the German tax bias proposed by Hafner and Wallmeier [2001] is described, design density (and consequently liquidity issues) are discussed in section 3.6.1. Estimation of the IVS and frequently observed phenomena observed by estimated IVS are discussed in the sections 3.6 and 3.7. Note that since the EURIBOR yield curves will be analyzed in section 4.2.5 the yield curves are discussed here in deeper extent than necessarily for the purpose of IV calculations.

3.5.1 Interest Rates (Yield Curves)

One of the parameters in the Black-Scholes pricing formula (2.6) is the interest rate r. The theoretical, risk-free, interest rate r, is not directly observable on the financial market and is typically *approximated* by the observed products, like interbank reference rates or bonds. It will be shown, that in practice, the interest rate is not constant, but depends on the time-to-maturity. The interest rate as a function of time-to-maturity – $r(\tau)$ is referred to as *yield curve*. In the next paragraph an yield curve based on the "EURIBOR" interest rate is introduced.

EURIBOR yield curve:
The EURIBOR (Euro Interbank Offered Rate) is a reference interest rate for interbank deposits in the euro zone. The EURIBOR is determined as a 15% trimmed average of the interest rates contributed by the "Panel banks – banks with the highest volume of business in the euro zone money market". (Cited from www.euribor.org, status at June, 10th 2006. For the complete list of panel banks, see www.euribor.org\html\content\panelbanks.html).

EURIBOR is determined for deposits with maturities 1 weeks, 2 weeks, 3 weeks, 1 month, 2 month, ... 12 months, see also section 3.6.1. An example for June 2nd, 2006 is given in the table 3.1, source www.euribor.org. The

maturity	rate	maturity	rate	maturity	rate
1 Week	2.612	2 Week	2.691	3 Week	2.764
1 Months	2.823	2 Months	2.904	3 Months	2.950
4 Months	3.019	5 Months	3.083	6 Months	3.126
7 Months	3.177	8 Months	3.224	9 Months	3.261
10 Months	3.302	11 Months	3.335	12 Months	3.368

Table 3.1: EURIBOR interest rates, June 2nd, 2006

interest rates in the table 3.1 obviously depend on the time-to-maturity. A common assumption, see Hull [2003], that simplifies matters is to assume that yield curve between the EURIBOR rates on the observable times-to-maturities is linear and is constant after the last observed value (12 months).

The approach of estimating the yield curve $r_t(\tau)$ (at the time t) by linearly interpolated EURIBOR quotes is popular in IV analysis, see Hafner [2004], Brunner and Hafner [2003] or Fengler et al. [2005]. Note that the EURIBOR r_{EUR} is yearly compounded interest rate, whereas the interest rate r used in BS formula (2.6) is a continuously compounded interest rate hence the transformation $r = \log(1 + r_{EUR})$ has to be performed. More precisely EURIBOR is quoted on an actual/360 day-accounting convention basis, see Hull [2003] for details.

The linear interpolation of the EURIBOR interest rate seems to be sufficient for the calculation of the implied volatilities, since the most frequently traded (most liquid) options have maturities smaller than 12 months. Yield curves, however can also be used for pricing and hedging the interest rates derivatives, and larger time-to-maturities may be considered, hence a natural question arises: how to enlarge the yield curve based on EURIBOR (EURIBOR yield curve).

The left boundary of the EURIBOR yield curve can be enlarged by using

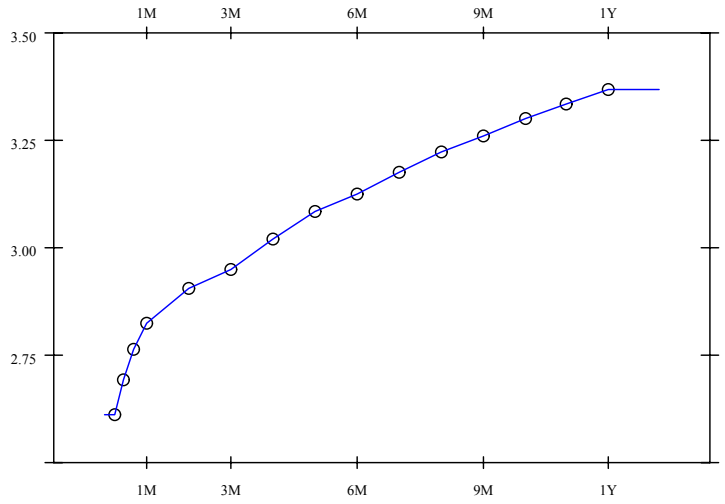

Figure 3.1: Euribor interest rate (points) and linearly interpolated yield curve (blue line) June 02, 2006.

the Euro Over Night Index Average (EONIA). EONIA index is the "effective overnight reference rate for the euro. It is computed as a weighted average of all overnight unsecured lending transactions undertaken in the interbank market, initiated within the euro area by the contributing banks" (cited from www.euribor.org, status at June, 10th 2006).

For maturities beyond 1 year the situation is more complicated – the money market does not contain the information about the yield curve for maturity higher than 1 year. Thus the right part needs to be constructed from different markets: Bonds market or derivative market (Futures or Interest Rate Swaps markets).

Basically, from the theoretical point of view, the easiest way of extracting the yield curve is by observing the implicit yields of the *zero coupon bonds*. A zero coupon bond is a financial instrument that simply pays its *nominal value* Z at the given expiration day T. Bond has a current market value P. Consider a zero-coupon bond with maturity in n-years and with nominal value Z then the corresponding yield *(zero yield)* is $[(Z/P)^{1/n} - 1]$. Since the zero bonds observed on the market have different maturities, the resulting yield curve is estimated using standard nonparametric regression, e.g. local polynomial estimates, see Benko [2002] among others.

Since the zero coupon bonds are usually not liquid enough, *coupon bearing bonds* are used to calculate the zero yields. A coupon bond is a bond

that yields a nominal value at the expiration day and extra-payments in predetermined time points in between (so called coupons) quoted in coupon rates. The method of subsequent calculation of zero rates from the cash-flows of different coupon bonds is known as *bootstrap*, for details see Hull [2003]. Note, that this bootstrap method has nothing to do with the well known resampling method in mathematical statistics.

The next popular interest rate derivatives, used for recovering the yield curve are the interest rate swaps. The *swaps* are contracts where two parties (e.g. company 1 and company 2) agree to exchange the future cash-flows. The *interest rate swaps* are contracts where one of the parties (e.g. company 1) agrees to pay to the counterpart cash flows equal to the fixed interest rate and obtain the cash flows corresponding to the floating interest rate, (vice versa the company 1 receive the "fixed interest rate" and pays the floating rates).

The floating interest rate is typically based on the EURIBOR (in EURO zone) or its conterpart London Inter Bank Offered Rate – LIBOR (in UK and US market), e.g. the floating rate is set up as 6 months EURIBOR + 0.5% p.a. Interest rate swaps are OTC products and a financial intermediary (as a bank) is involved and makes the market for swaps – negotiate the agreement individually with both companies. Of course the bank has to earn money and if for example the fixed rate of a swap is received by the bank, than this is higher than the fixed rate of a swap that the bank pays. The average of this bid-ask spread is called *swap rate* and is published by the financial data vendors. As argued in Hull [2003] a swap is a difference of fix rate and floating rate bonds, the collection of swaps define a collection of bonds, and the zero yields can then be calculated using bootstrap method.

The zero yield curve based on the swap contracts is plotted in the figure 3.2. The points are input data taken from Ecowin Reuters database. The blue line is linear interpolation of observed points.

In fact the figures 3.1 and 3.2 show a yield curve that is upward slopping - there are also example where the yield curve is downward slopping or even humped. A yield curve constant over maturities is often referred to as a flat yield curve in financial literature, an increasing yield curve is called normally shaped and decreasing an inverse yield curve.

A natural question arises: Why the yield curve is not flat ?
First, one may argue that, in theory, we are considering a *risk-free* interest rate and an interest rate observed on the market is never completely risk-free. On the other hand, significant yield curve effect can be observed also by the yield curve recovered from government bonds, where the risk is almost zero, see Benko [2002] among others. Hence different theoretical arguments for this phenomenon arose. Two arguments taken from Hull [2003] are:

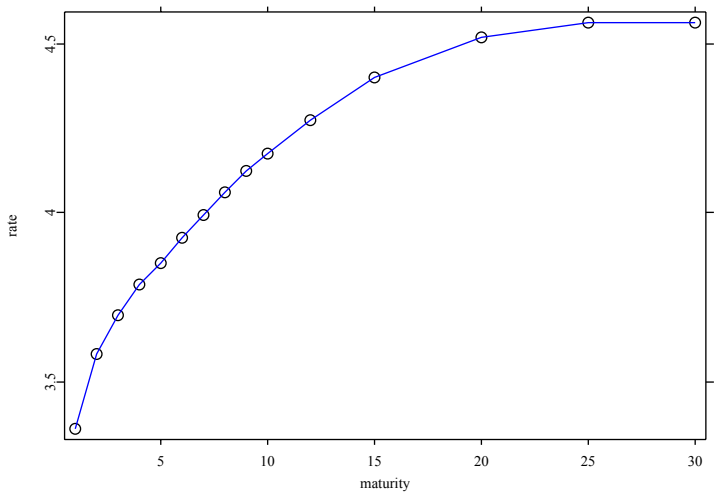

Figure 3.2: Zero yields - swap based (points) and linearly interpolated yield curve (blue line) June 02, 2006.

- Market Segmentation: Different maturities reflects different markets and the observed yields are determined by supply and demand of these specific markets.

- Expectation Theory: The interest rate for longer maturities reflects the expectation of the market about the future short term interest rates. This can be formalized by consideration of the *forward interest rates* – interest rates applied for future time period. For example a *forward rate* for time period between T_1 and T_2 ($T_2 > T_1$) is specified by the zero rates $r(T_1)$ and $r(T_2)$ by:

$$r_F(T_2, T_1) = \frac{r(T_2)T_2 - r(T_1)T_1}{T_2 - T_1} = r(T_2) + (r(T_2) - r(T_1))\frac{T_1}{T_2 - T_1}. \tag{3.30}$$

  Instantaneous forward rate (a interest rate valid for short maturities in future) is obtained as the limit of (3.30):

$$r_F(T) = r + T\frac{\partial R}{\partial T}. \tag{3.31}$$

  The equations (3.30) and (3.31) give also no-arbitrage conditions on the yield curve - the forward rates implied by the zero yields curve have to be non-negative. In order to keep the notation simple, the yield curve is here understood as a function of the maturity date.

The importance of forward rates is "justified" in practice by the existence of *forward-rate agreements* where the parties agree on an interest rate that applies in future time for given loan.

Again there are different types of models for yield curves, one stream based on the diffusion models (similar to those introduced for option prices in chapter 2). Generally these are called short rate models. Prominent models are Vasicek, Cox-Ingersoll-Ross or Hull-White models, see Zagst [2002], section 4.5.1 among others.

Next important stream of stochastic based literature are the LIBOR Market Models, proposed for real market rates, see Zagst [2002], section 4.7 among others, for discussion and bibliographic overview.

Another approach proposed by Nelson and Siegel [1987] is based on the factor model with fixed and known factor functions representing level, slope and curvature of the yield curves.

3.5.2 Option Prices

As already argued the implied volatility introduced in section 2.3 is essentially calculated from the option prices by inverting the Black-Scholes Formula (2.6). In this section we will discuss the phenomena connected to this process.

First of all we focus on the most liquid options traded on the German-Swiss exchange (EUREX) – the options on German stock index DAX 30. DAX 30 is a capital weighted performance index based on the 30 largest German companies traded on the exchange (large in terms of capitalization and liquidity). The dividends of the companies included in DAX (DAX companies) are reinvested into the index.

The options on DAX (ODAX) are European styled options (Put and Calls). For the options traded on the exchange the expiry days are standardized – the expiration of the option is fixed by the expiration month. By convention the exact expiration day is the third Friday of the contract month (if it is a trading day, if not, then the expiration day is the closest trading day before).

At the given trading day the following contract months are available: Up to 60 months: The three nearest successive calendar months, the three following months of the March, June, September and December (quarter year cycle), the four following months of the June and December (semi-annual cycle) the altogether seven following annual months of the December cycle. (Source: www.eurexchange.com, status Sep, 27, 2006). Note that the options with maturity longer than 2 years are traded since March 2003. According to the product description, the exchange offers at least 7 (3 ITM,

3 OTM and ATM) different strikes for each available time-to-maturity, for maturity shorter than 24 months and at least 5 (2 ITM, 2 OTM and ATM) for longer maturities. However typically more strikes are available - the smaller difference in the subsequent strikes is 50 points. For definition of at-the-money (ATM), in-the-money (ITM) and out-of-the-money (OTM), see section 3.6.1.

The Eurex Information Services offers two basic types of data, that includes options and futures, first the "Intra-day" data, that records the characteristics of each contract on the exchange. Currently it consists of product id – relevant for our case are ODAX (options) and FDAX (futures), call put flag – identification for put, call and future trades, expiration month and year, exercise price (K), time of trade (t) year, month, day, hour, minute, second, centiseconds and miliseconds, price of trade, trade size, moreover it consists of variables not addressed in this study: version (used for capital adjustments), currency and trade type.

The Daily data (or Settlement data) are data valid for a given day and consist of product id, call put flag, expiration month and year, exercise price, (fact) year, month and day, opening price, highest price, lowest price, "settlement" price, volume of traded contracts, version and open interest (number of open positions). The settlement data could be basically understood as closing data, however it can happen that for some option type the last traded option for a given day is traded on a far earlier time point than the others, in this case the settlement price is established by exchange, the reason for this is for example the aim to obtain data fulfilling some natural (e.g. no-arbitrage) conditions, see also section 3.8.

Next interesting data basis is the order book data (or order book), where not only the settled contracts are recorded but also the whole history of the orders from both supply and demand part.

3.5.3 Spot Prices

Generally speaking the spot price for the underlying can be obtained from the stock exchange market (if the underlying is a single stock) or from the index vendors (for the DAX index considered here, the vendor is the EUREX exchange). Nevertheless it is a common practice to recover spot prices from derivative markets (for underlying that are liquid on these markets), see Hafner and Wallmeier [2001], Fengler [2005b]. This seems to be reasonable, since although the spot and derivative markets are naturally connected, some differences in the market mechanisms can affect the calculation. There is also a practical advantage since the procedure uses only one data source.

In a frictionless financial market for asset without dividend payments, is

the no-arbitrage price of a future contract (see also discussion in the section 2.4.2) given by

$$S_t = e^{-r_t(\tau_F)\tau_F} F_{t,\tau_F} \ , \tag{3.32}$$

where the τ_F denotes the time-to-maturity of the future contract, the S_t is the spot (index) price and $r_t(\tau_F)$ is the interest rate valid for the time-point t and time-to-maturity τ. Typically the most liquid future contract $F_{t,\tau}$ is chosen in practice. For the calculation of the IV-intra-day data in the MD*Base/FEDC IV database, considered here, the most liquid future (w.r.t. trading volumes at the given day) has been used.

Calculation for the daily (closing) data is straight forward. If the intra-day (contract based) data are considered, one complication occurs: there is no guarantee that a future is traded in the same time as the corresponding option. A future price of nearest available future contract within the one minute interval is used for calculation of the corresponding underlying price.

3.5.4 Adjusted Spot – Data Correction Scheme

The DAX index is a capital weighted performance index, i.e. dividends (after corporate tax) are considered as reinvested into the index. Therefore, dividends should have no impact on the index options. However, when only the interest rate discounted futures price is used to recover IVs by inverting the BS formula, IVs of calls and puts can differ significantly (as already discussed in a frictionless market without dividends, the put and call prices should lead to IV that coincide). This section is based on the procedure proposed by Hafner and Wallmeier [2001], compare also with Fengler [2005b].

An topic that influences spot price calculations is corporate tax applied to the dividends. In the former taxation legislative in Germany applied from 1977 to 2001, a gross dividend paid by a company was subject to a corporate tax ("Körperschaftssteuer") paid directly to the company for domestic shareholders (income tax residents). This tax was taken into account in the personal income tax of the shareholders in so called "Anrechnungsverfahren" (tax-credit procedure). However this procedure did not apply to the foreign investors. In oder to correct this "German tax bias", the option data has undergone a preparation scheme proposed by Hafner and Wallmeier [2001] that is described below.

Note, that in 2000 an important set of changes in income taxation were introduced in Germany (Steuersenkungsgesetz, BGBl. Teil I, Nr. 46 dating from November 26, 2000) with transition period in 2001 and coming into full force in 2002. Since 2002, the taxes paid on corporate tax can no longer be used as a tax-credit by domestic shareholders. Instead, 50% of the

distributed dividends are taxed at the personal income tax ("Halbeinkünfteverfahren"), the other 50% are not liable to any further taxation. In the FEDC database the correction scheme have been applied to all data due to the consistence reason. However due to taxation scheme changes in 2000, the correction procedure may no longer be necessary for the DAX index option data beginning from 2002, this question might be an interesting topic of further investigations.

Hafner-Wallmeier Correction: Hafner and Wallmeier [2001] argue that the marginal investor's individual tax scheme is different from the one actually assumed in the calculation of the DAX index, see discussion above. Differences occur between the domestic shareholders (German tax residents) and foreign investors, as already discussed, as well as between German tax residents with different personal income tax rates (since Germany has a progressive personal income tax scheme).

Consequently, the net dividend for different investors can be higher or lower than the one used for the index computation. The discrepancy, which Hafner and Wallmeier [2001] call *difference dividend*, has the same impact as a dividend payment for an option – affects the price and consequently the IVs. Denote by $\Delta D_{t,\tau}$ value of this difference dividends between t and T at the time $T = t + \tau$. Consider the dividend adjusted futures pricing formula:

$$F_{t,\tau_F} = e^{r_t(\tau_F)\tau_F} S_t - \Delta D_{t,\tau_F} \ , \tag{3.33}$$

and the dividend adjusted put-call parity:

$$C_t(K,\tau_H) - P_t(K,\tau_H) = S_t - \Delta D_{t,\tau_H} e^{-r_t(\tau_H)\tau_H} - e^{-r_t(\tau_H)\tau_H} K \ , \tag{3.34}$$

with τ_H denoting the time-to-maturity date of the options. Inserting equation (3.33) into (3.34) yields

$$C_t(K,\tau_H) - P_t(K,\tau_H) = F_{t,\tau_F} e^{-r_t(\tau_F)\tau_F} + \Delta D_{t,\tau_H,\tau_F} - e^{-r_t(\tau_H)\tau_H} K \ , \tag{3.35}$$

where $\Delta D_{t,\tau_H,\tau_F} \stackrel{\text{def}}{=} \Delta D_{t,\tau_F} e^{-r_t(\tau_F)\tau_F} - \Delta D_{t,T_H} e^{-r_t(\tau_H)\tau_H}$ is the desired difference dividend.

The "adjusted" index level

$$\tilde{S}_t = F_{t,\tau_F} e^{-r_t(\tau_F)\tau_F} + \Delta D_{t,\tau_H,\tau_F} \tag{3.36}$$

is that index level, which "matches" the put and call IVs exactly to the same levels.

Of course the difference dividend $\Delta \hat{D}_{t,\tau_H,\tau_F}$ has to be estimated, using (3.35), since contract based data are used. The pairs of puts and calls of

the strikes and same maturity are matched (to each put a call trade) if the difference between their trading times is smaller than five minutes. The final value of $\Delta\hat{D}_{t,\tau_H,\tau_F}$ is estimated by the median of all $\Delta D_{t,\tau_H,\tau_F}$ for a given maturity at day t, the median is used in order to guarantee the robustness of the approach. The IVs are recovered by inverting the BS formula using the corrected index value

$$\tilde{S}_t = F_{t,\tau_F} e^{-r_t(\tau_F)\tau_F} + \Delta\hat{D}_{t,\tau_H,\tau_F}.$$

Summarizing, the calculation of the implied volatility proceed as follows:

1. Collecting raw data: ODAX, FDAX Prices (Source: EUREX), EURIBOR Interest rates from 1,3,6 months and 1 year (Source *Thomson Financial Datastream* , *Ecowin* FIBOR interest rates were used for period between 1995 and 1999) and Expiration dates obtained from trading calendars published by EUREX.

2. Matching the raw data – for each ODAX and FDAX contract an expiration date, time-to-maturity and an interest rate corresponding to this time-to-maturity obtained by linear interpolation of the raw interest rates.

3. Hafner-Wallmeier correction as described in the section 3.5.4 is performed – corrected spot $\tilde{S}_t$ is used as a Spot in the BS Formula.

4. IV is calculated by Newton-Raphson iterative procedure.

First two steps are performed in the Oracle database system using SQL plus language, the steps 3 and 4 in statistical software XploRe.

As already mentioned, by the HW correction, some part of the data – calls and puts that could not be matched into the put-call parity (3.34) are discarded. For the ODAX option-prices and all years starting from January 1995 to June 2005 roughly 95% of all option contracts an IVs can be calculated. Table 3.2 shows the comparison of the number of the Put and Call options in the option-price data set and the number of contract after the Hafner-Walmeier procedure. It will be showed in the section 3.6.1 that the trading intensity is decreasing with the time-to-maturity. Consequently, more option prices are discarded for higher maturities, e.g. for year 2004 IVs were calculated, using the HW procedure as described above, from more than 96% of option prices, for both puts and calls, with time-to-maturity shorter than 1 year whereas only 25% of the option with longer maturity.

Year	Puts Raw	Puts HW	Calls Raw	Calls HW
1995	243	221	263	239
1996	248	236	254	242
1997	361	336	348	325
1998	431	413	385	369
1999	467	450	458	443
2000	511	488	455	436
2001	504	489	462	446
2002	450	433	429	413
2003	442	424	415	400
2004	810	774	736	704
2005*	379	366	356	344

Table 3.2: Comparison of number of contracts (in thousands) in the option data against the number of IVs after Hafner-Walmeier procedure. For years 1995 – 2005, (for 2005 only first half year is used).

3.6 Estimating the IVS

The IV has been introduced in the theoretical setup in the chapter 2. The set of assumptions on the efficiency of the financial market yields for a fixed time point an unique IV for given K, τ – $\sigma(K_i, \tau_i)$. However "nothing is perfect in practice". The IVs obtained from the option prices observed on the market may be subject to noise.

Denoting IVs observed on the market (more precisely calculated from the option prices observed on the market) at K_i and τ_i by $\tilde{\sigma}_i$ and the corresponding "true" IV function by $\sigma(K_i, \tau_i)$, $i = 1, \ldots, T$ we assume the following model:

$$\tilde{\sigma}_i = \sigma(K_i, \tau_i) + \varepsilon_i \ , \tag{3.37}$$

The noise occurs due to market micro-structures effects, such as bid-ask spreads, discrete ticks in prices and quotes, non-synchronous trading, effects due to the auction mechanism itself, or simply due to misprints, for a detailed analysis of errors in IV data we refer to Hentschel [2003].

Although there are some attempts to model IVs by parametric techniques, see Brockhaus et al. [2000] chapter 2 among others, the nonparametric estimation techniques presented in section 3.1 are natural candidates, since only smoothness of the IVS is assumed. Techniques based on local polynomial estimation are popularly employed, see Shimko [1993], Fengler et al. [2003] and Cont and da Fonseca [2002] for such applications to estimation of the IV

surface. In the following section 3.7 some phenomena and typical features of IVS are illustrated. In the sections 3.6.1 and 3.7 standard nonparametric techniques are employed, whereas the section 3.8 discusses no-arbitrage estimation of the IVS, originally presented in Benko et al. [2006a].

3.6.1 Liquidity Issues, Moneyness and Time to maturity

In this section we discuss the empirical findings by investigating the shape of the "observational design" of the IVs (K, τ), as in definition 3.1, of IVs based on the ODAX options traded at EUREX. The official product description that gives basic setup of this investigation is introduced in the section 3.5.2. We begin with the transformation of the strike direction of the design:

Moneyness

A call option has payoff $\max(S_T - K, O)$, i.e. has positive payoff for $S_T > K$, vice versa for a put option. Clearly the price of an option depends heavily on the current level of the underlying (S_t). An option is said to be *in-the-money* (ITM) at t if it yields positive payoff by exercising at t, i.e. if $S_t > K$ for call options and $S_t < K$ for put options. On the other hand, an option is said to be *out-of-the-money* (OTM) if it yields "negative payoff", i.e. $S_t < K$ for call options and $S_t > K$ for put options. The point $S_t = K$ is called *at-the-money* (ATM). The indicator of ATM/OTM/ITM and its quantification is refereed to as *moneyness*. Different *moneyness measures* have been proposed. Hafner [2004] defines moneyness as a function $m(t, S_t, K, T, r)$ that is increasing in K.

Simple moneyness based on the introduction above is defined by

$$\kappa_{spot} = \frac{K}{S_t}$$

and is also called *stock price moneyness* or *simple moneyness*, see also Fengler [2005b]. It is often used by traders. Using the stock price moneyness, a call option is OTM if $\kappa_{spot} > 1$, ATM if $\kappa_{spot} = 1$ and ITM $\kappa_{spot} < 1$, vice versa for a put option.

In the previous section we have already introduced the so called *futures moneyness* defined as

$$\kappa = \frac{K}{F_{t,\tau}} \tag{3.38}$$

where $F_{t,\tau} = S_t e^{r\tau}$, see formula (2.21). Thinking in the future moneyness an "ATM" point is defined by $\kappa = 1$ and OTM, ITM in the same way as for κ_{spot}.

The call-equivalent BS Delta of the option, defined in section 2.1, is also used as measure of moneyness, see Hafner and Wallmeier [2001]. Here a

call option is ATM for Delta equal 0.5, ITM and OTM for Delta above and below 0.5 respectively. Connection between Delta and κ as well as the difference between ATM defined in terms of κ_{spot} and κ are briefly discussed in Hafner [2004]. Deeper discussion of the theoretical aspects of the moneyness reparametrization can be found in Renault [1997].

Popular are also transformations of futures moneyness: *log-futures moneyness* $\log(\kappa)$, see Hafner [2004]. The standardization of the expiration time T is typically done by already used time-to-maturity τ. For the purpose of this paper, we use τ measured in years in (actual/360 day count standard). Moneyness measures combining K and τ together have been proposed (besides simple futures moneyness) by Natanberg [1994] as ratio of $\log \kappa$ and $\sqrt{\tau}$ or so called standardized moneyness, essentially obtained by the further division by the corresponding ATM IV proposed by Tompkins [1994].

The design in the time-to-maturity (τ) is affected by the market standardization, on the exchange – options with specific expiration dates are traded – the expiry date is a fixed day in given month, see section 3.5.2 for exact product description. As consequence, time varying and degenerated design in the time-to-maturity is observed. Figure 3.3 gives an example on randomly chosen day (May 22nd, 2003), on this day IVs with maturities 29, 57, 85, 120 and 211 days-to-maturity were observed.

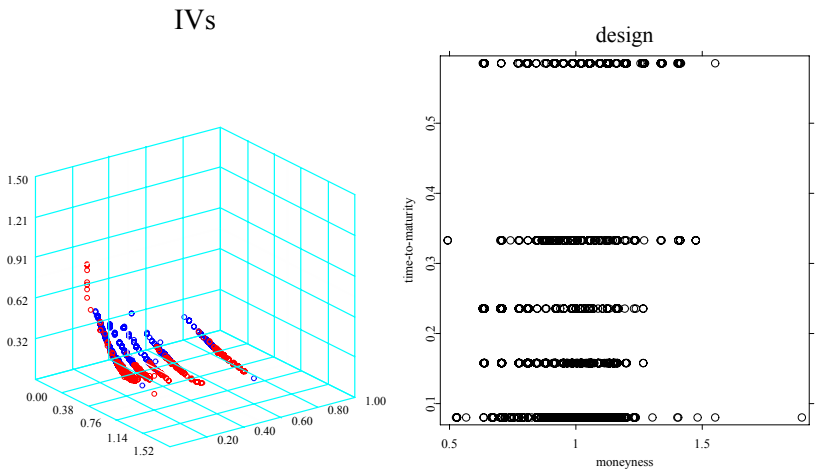

Figure 3.3: Left figure – IVs observed on May 22nd, 2003, blue points are IVs calculated from the Puts, red points from Calls and the corresponding design (right figure).

Tables 3.6.1 and 3.6.1 display the distribution of the future moneyness and time-to-maturity for the puts (table 3.6.1) and calls (table 3.6.1), for

κ/τ	1M	2M	3M	6M	1Y	2Y	2-4Y	$> 4Y$	cum.
$[0.0.5)$	2	46	22	246	398	62	6	0	782
$[0.5, 0.8)$	1662	3574	3912	6660	5938	460	34	0	22240
$[0.8, 0.9)$	21480	31344	19624	13042	6224	306	32	0	92052
$[0.9, 1)$	322882	123666	37488	26620	7154	682	58	2	518552
$[1, 1.1)$	91670	25236	6224	5430	2202	222	24	0	131008
$[1.1, 1.2]$	2300	604	366	688	486	100	2	0	4546
> 1.2	1662	438	594	1082	544	94	2	0	4416
cum	441658	184908	68230	53768	22946	1926	158	2	773596

Table 3.3: Cross table – IV design, Puts 2004

κ/τ	1M	2M	3M	6M	1Y	2Y	2-4Y	$> 4Y$	cum.
$[0.0.5)$	148	82	78	112	146	20	0	0	586
$[0.5, 0.8)$	1570	386	270	770	366	44	0	0	3406
$[0.8, 0.9)$	4148	1150	618	920	726	130	4	0	7696
$[0.9, 1)$	138256	31496	9538	7030	2878	250	72	2	189522
$[1, 1.1)$	269592	122296	35602	22988	6974	400	28	2	457882
$[1.1, 1.2]$	3836	8922	8758	11260	5592	250	10	0	38628
> 1.2	190	240	274	1492	3948	470	52	2	6668
cum	417740	164572	55138	44572	20630	1564	166	6	704388

Table 3.4: Cross table – IV design, Calls 2004

IVs based on the DAX option prices traded on the EUREX after Hafner-Wallmeier Correction presented in the previous section. Time-to-maturity is discretized in the following way: 1M interval $[0, 30/360)$, 2M interval $[30/360, 60/360)$, 3M interval $[60/360, 90/360)$, 6M interval $[90/360,\ 0.5)$, 1Y interval $[0.5, 1)$, 2Y interval $[1, 2)$ 2-4Y interval $[2, 4)$ and >4Y interval for $[4, \infty]$, all measured in years.

It can be seen that most of the IVs are available for $\kappa \in [0.8, 1.2]$ and for short maturities. Figures 3.4 and 3.5 illustrate the densities of (κ, τ) for $\kappa \in [0.75, 1.25]$ and $\tau \in [0, 0.3]$ and display contour plot of the KDE estimators using product Quartic kernel. Bandwidths are chosen by Scott's rule, see Scott [1992].

From the presented tables, it can be easily concluded that the majority of the IVs are available for short maturities. For maturities longer than 1 year only small number of observations is available. In the moneyness direction we can observe that the IVs are mainly available and consequently the options are traded much more frequently close to the ATM points. The design density of IVs obtained from put options is left skewed and calls right skewed. For $\kappa < 1$ the puts are more liquid than calls, vice versa for $\kappa > 1$ and OTM options are traded more frequently than ITM. To illustrate these phenomena, KDE estimate of the marginal density w.r.t. κ with Quartic

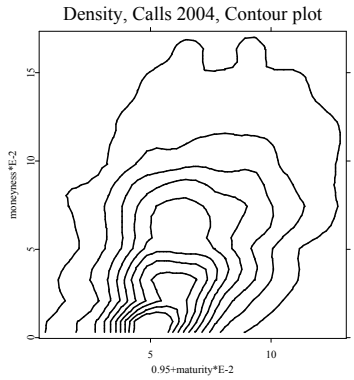

Figure 3.4: Contour plot, the density of IVs w.r.t. κ and τ for Calls, year 2004.

kernel and bandwidth chosen by Silverman rule is displayed on the figure 3.6.

As consequence, there is different liquidity level for put and call options for different areas of the future moneyness, connected liquidity risk may affect the price of the options and IVs of the Puts and Calls, although theoretically equal, may differ. This liquidity effect is more pronounced for far from ATM.

The analysis is illustrated by the currently last complete year in the FEDC EUREX IV database (2004). Similar results for the years 1995 – 2004 and first half of the year 2005 have been obtained, for illustration the contour plots of the design densities can be found in Appendix B, figures B.4 – B.7.

3.7 Implied Volatility Surface - Empirical Findings

In this section static features typical for IVS are discussed in addition to the comments on design of the IVs given in the section 3.6.1. The analysis is based on the IVs ODAX starting from 1995 to 2005 (for 2005 only first half year was available). Figure 3.7 displays the average of daily local linear estimates of the IVs for year 2004. The bandwidth has been chosen individually (for each day) as the smallest bandwidth that guarantees that the estimates are well defined for $\kappa \in [0.9, 1.05]$ and $\tau \in [0, 0.25]$. This choice is justified since we are estimating the mean and the averaging is decreasing variance but not bias.

The average IVS for years 1995 up to 2005 (for 2005 only first half year were available) are displayed in appendix, figures B.2 and B.3, standard

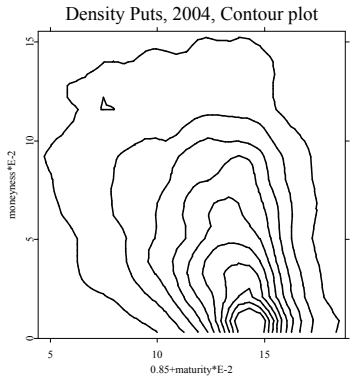

Figure 3.5: Contour plot, the Density of IVs w.r.t. κ and τ for Puts, year 2004.

deviations for ATM are displayed in the figure B.1. It can be seen that the curvature of the IVS is more pronounced for shorter maturities and more flat for longer maturities.

The smile typically achieves its minimum close to the ATM. The standard deviation (volatility of volatility) of the local linear IVs estimations are typically decreasing in τ, see figure B.1 in appendix. Note that the IVs exhibits strong leverage effects – strong negative correlation between returns of the IV and returns of the underlying (DAX), see also figure 2.2 and the connected discussion in the section 2.5.

In the next section we will discuss the problems of estimation of the IV functions and surfaces in much more deeper, focus is on the estimation of the IVS and no-arbitrage. Sections 3.8 follows closely arguments of Benko et al. [2006a], where the method was originally proposed.

3.8 Fitting the IVS and No-arbitrage

The IV function and the SPD are naturally connected. This motivates the estimation technique based on the combination of these two concepts. The section 3.8.1 presents such a method for estimating the IV function for a given maturity τ observed on the given day, section 3.8.2 is devoted to estimation of the IV surface. Both methods are applied to the ODAX IVs.

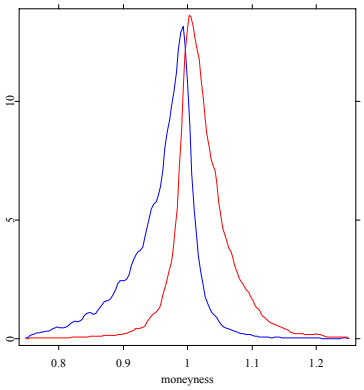

Figure 3.6: Plot of marginal densities w.r.t. κ, (finely dashed) red line for Calls and (solid) blue line for Puts, year 2004.

3.8.1 Estimating the IV Smile for Fixed Maturity

In this section, we focus on estimating the IV function from a set of observed option prices with fixed time-to-maturity τ. Since, in this section τ is fixed, we may simplify notation in (3.37) by setting $\sigma(K_i) \stackrel{\text{def}}{=} \sigma(K_i, \tau_i)$ for $i = 1, \ldots, T_\tau$, T_τ denotes the number of observed IVs with maturity τ. Then (3.37) simplifies to

$$\tilde{\sigma}_i = \sigma(K_i) + \varepsilon_i. \tag{3.39}$$

In the figure 3.8 we present IVs calculated from the daily option prices on February 2, 2006 for the maturity of 15 days. As can be seen the IVs are quite rough and specially the boundary regions (strike below 5000 and above 6000) two IV values are observed for the same strike (corresponding to puts and calls). By the put-call parity, these IVs must coincide.

The discrepancy observed in IV in figure 3.8 is modeled via the error term in (3.39). As already outlined, our aim is to combine the regression model (3.39) with the SPD in (2.12). According to (2.12) the SPD is a function of the IV function and its first and second derivative. We therefore propose a local quadratic estimator that automatically provides an estimate of the IV function and its derivatives, see Fan and Gijbels [1996]. Referring to the section 3.2, the local quadratic estimator $\hat{\sigma}(K)$ of the regression function $\sigma(K)$ in the point K is defined by α_0 – minimizer of the following local least squares criterion:

$$\min_{\alpha_0,\alpha_1,\alpha_2} \sum_{i=1}^{T_\tau} \left\{\tilde{\sigma}_i - \alpha_0 - \alpha_1(K_i - K) - \alpha_2(K_i - K)^2\right\}^2 \mathcal{K}_h(K - K_i), \tag{3.40}$$

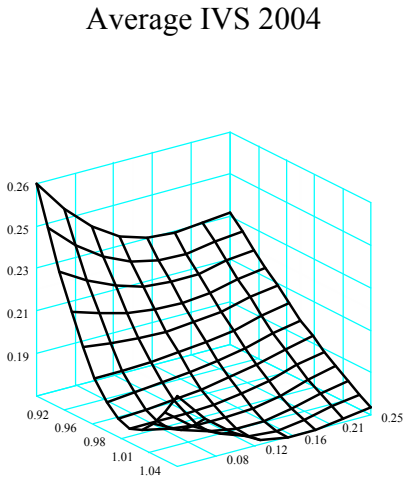

Figure 3.7: Yearly averaged smoothed (by local linear estimate) for whole year 2004.

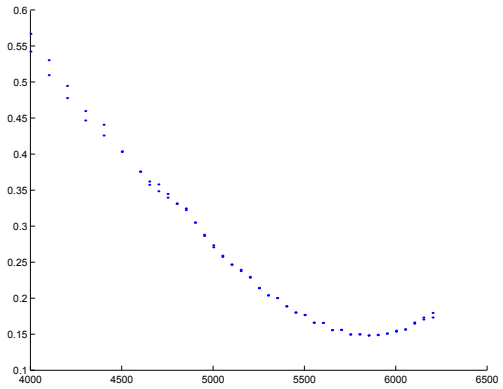

Figure 3.8: Daily IVs, ODAX on February 2, 2006 with maturity 15 days, horizontal axis is the strike level, vertical axis the volatility.

where $\mathcal{K}_h(K-K_i) \stackrel{\text{def}}{=} \frac{1}{h}\mathcal{K}\left(\frac{K-K_i}{h}\right)$ Comparing (3.40) with the Taylor expansion of σ yields

$$\alpha_0 = \hat{\sigma}(K_i),\ \alpha_1 = \hat{\sigma}'(K_i),\ 2\alpha_2 = \hat{\sigma}''(K_i), \tag{3.41}$$

which makes the estimation of the regression function and its first two derivatives ($\hat{\sigma}'$ and $\hat{\sigma}''$ respectively) possible. In order to take the non-negativity of the SPD into account, we need to perform (3.40) under the condition $q_{t,S_T} \geq 0$, on the entire support. Since we consider only one time point t we will ease the notation in this section to $S_t = S$, and $q_{t,S_T}(K,\tau)$ to $q(K,\tau)$. For the fixed point (K,τ) and by plugging (3.41) into (3.40) we can rewrite $d_1 = \frac{\ln(S/K)+(r+0.5(\alpha_0)^2)\tau}{\alpha_0\sqrt{\tau}}$, $d_2 = d_1 - \alpha_0\sqrt{\tau}$ and the SPD can be estimated at the point (K,τ) by

$$\hat{q}(K,\tau) = F\sqrt{\tau}\phi(d_1)\left\{\frac{1}{K^2\alpha_0\tau} + \frac{2d_1}{K\alpha_0\sqrt{\tau}}\alpha_1 + \frac{d_1 d_2}{\alpha_0}(\alpha_1)^2 + 2\alpha_2\right\}, \tag{3.42}$$

where $F = Se^{r\tau}$. Summarizing, the optimization problem can be written as:

$$\min_{\alpha_0,\alpha_1,\alpha_2} \sum_{i=1}^{T_\tau} \left\{\tilde{\sigma}_i - \alpha_0 - \alpha_1(K_i-K) - \alpha_2(K_i-K)^2\right\}^2 \mathcal{K}_h(K-K_i) \tag{3.43}$$

subject to $F\sqrt{\tau}\phi(d_1)\left\{\frac{1}{K^2\alpha_0\tau} + \frac{2d_1}{K\alpha_0\sqrt{\tau}}\alpha_1 + \frac{d_1d_2}{\alpha_0}(\alpha_1)^2 + 2\alpha_2\right\} \geq 0$ where $d_1 = \frac{\ln(S/K)+(r+0.5(\alpha_0)^2)\tau}{\alpha_0\sqrt{\tau}}$, $d_2 = d_1 - \alpha_0\sqrt{\tau}$. As already mentioned, this leads to a nonlinear optimization problem. All computations of parameters α_0, α_1, α_2 were done in GAMS 22.0 - solver MINOS. For an overview on the nonlinear optimization, see Bertsekas [1999] among others.

The constrained estimate, and the corresponding SPD for the dataset using the Epanechnikov kernel and $h = 200$ are displayed on the figure 3.9. The smoothing parameter has been chosen by keeping bias small on the one hand and guaranteeing enough data for each point κ where the estimate (3.43) was constructed. More sophisticated choice of the parameter h seems to be possible, by using the standard cross-validation argument, see e.g. Härdle [1990]. Adaptive methods proposed recently by Spokoiny [2006] might by also employed in smoothing parameter selection, under some further conditions. The confidence intervals for the estimated IV and the SPD can be constructed by using classical idea of wild residual bootstrap, see Härdle [1990] among others.

Let us consider now the more complicated situation using the intra-day data. The IVs are calculated for each realized trade on the EUREX exchange.

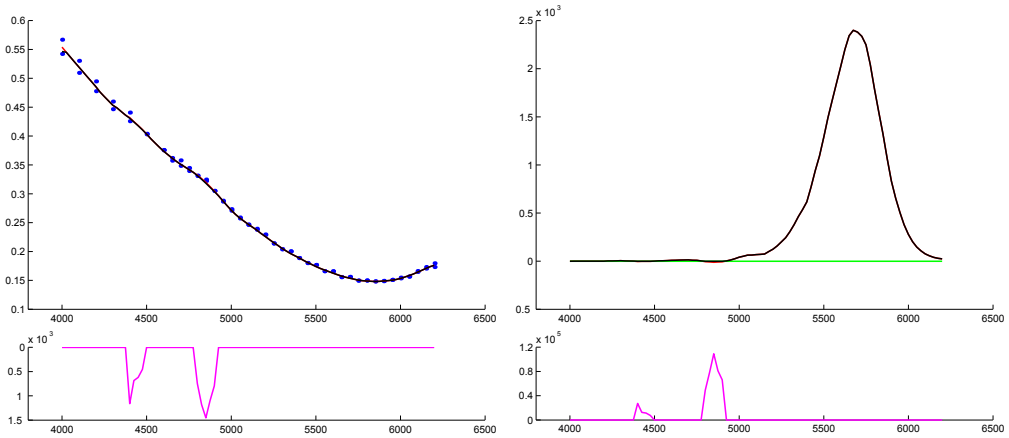

Figure 3.9: Left figure: Smoothed IV function. Black line (constrained) red line (unconstrained) and corresponding SPDs–black line (constrained) and red line (unconstrained), daily data on February 2, 2006, horizontal axis is the strike level. The lower figures illustrate the difference between constrained and unconstrained smoothing.

The crucial difference in comparison to the daily data is that the underlying DAX prices that are used in the calculation of the IV are not constant over time. In order to standardize the IV w.r.t. to underlying stock (in our case DAX index), we express the IV as a function of *futures moneyness* ($\kappa = \frac{K}{F}$). Using this standardization, the local quadratic estimate $\hat{\sigma}(\kappa)$ of $\sigma(\kappa)$ is given by α_0 that solves:

$$\min_{\alpha_0,\alpha_1,\alpha_2} \sum_{i=1}^{T_\tau} \left\{\tilde{\sigma}_i - \alpha_0 - \alpha_1(\kappa_i - \kappa) - \alpha_2(\kappa_i - \kappa)^2\right\}^2 \mathcal{K}_h(\kappa - \kappa_i). \tag{3.44}$$

Next, from the definition of futures moneyness we obtain: $K = F\kappa$, $\frac{\partial K}{\partial \kappa} = F$, $\frac{\partial \sigma}{\partial K} = \frac{\partial \sigma}{\partial \kappa}\frac{1}{F}$, $\frac{\partial^2 \sigma}{\partial^2 K} = \frac{1}{F^2}\frac{\partial^2 \sigma}{\partial^2 \kappa}$. After some straightforward calculations we obtain

$$d_1 = \frac{-\ln(\kappa e^{r\tau}) + (r + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}} = \frac{\sigma^2\tau/2 - \ln(\kappa)}{\sigma\sqrt{\tau}},\ d_2 = d_1 - \sigma\sqrt{\tau}.$$

Finally we obtain the SPD expressed as a function of κ (note that after analytical calculations the SPD needs to be rescaled in order to have $\int q(\kappa)d\kappa = 1$):

$$q(\kappa,\tau) = \sqrt{\tau}\phi(d_1)\left\{\frac{1}{\kappa^2\sigma\tau} + \frac{2d_1}{\kappa\sigma\sqrt{\tau}}\frac{\partial\sigma}{\partial\kappa} + \frac{d_1 d_2}{\sigma}\left(\frac{\partial\sigma}{\partial\kappa}\right)^2 + \frac{\partial^2\sigma}{\partial\kappa^2}\right\}. \tag{3.45}$$

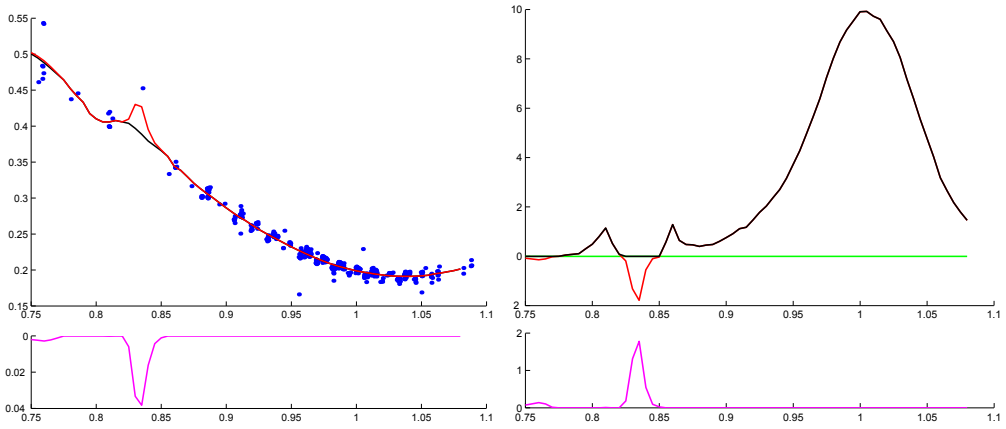

Figure 3.10: Left figure: Smoothed IV function. Black line (constrained) red line (unconstrained) and corresponding SPDs–black line (constrained) and red line (unconstrained), intra-day data, December 29, 2003, horizontal axis is moneyness level. The lower figures illustrate the difference between constrained and unconstrained smoothing.

Hence the analogue of (3.43) can be obtained by constraining (3.44) w.r.t. the corresponding non-negative SPD (3.45).

The left plot of the figure 3.10 shows the intra-day data (blue points) on December 29, 2003, red line is the constrained local quadratic smoother with Epanechnikov kernel and $h = 0.045$, the black line is the constrained local polynomial estimator with the same bandwidth.

Since S_t and F_t are not constant in intra-day data, daily average S of (S_t) and $F = Se^{r\tau}$ is used in (11). As a nice side-effect, we can see that the constrained estimator is more robust against outlyers. The corresponding SPDs are plotted in the right plot.

It should be noted that we are using one functional optimization criterion (3.43) for estimating the function (IV) and its first and second derivative simultaneously. As argued in Fan and Gijbels [1996] if we were interested in these functions separately, it could be advantageous to consider a separate objective function for each of these functions with different bandwidths or different order of the polynomial used in (3.43). However, the elegance of our approach is that all quantities needed for determining the SPD are obtained from (3.43) immediately.

The confidence intervals for the estimated IV and the SPD can be constructed by using classical idea of wild residual bootstrap, see Härdle [1990] among others.

Note that in this chapter we focus only on the positivity of the SPD function, the integral condition ($\int q_{t,S_T}(s,\tau)ds = 1$) might be considered as well in the same way, however since this condition is not a local but global condition, computationally this will be more involved. Moreover in this case the tails of the SPD that are outside of the observed implied volatility need to be considered, essentially these tails can be estimated simply by the cumulative distribution function $F_{Q_{t,S_T}}(x) \overset{\text{def}}{=} \int_0^x q_{t,S_T}(s,\tau)ds$ corresponding to the SPD (2.17) yields to:

$$F_{Q_{t,S_T}}(x) = e^{r\tau}\frac{\partial C_t(K,T)}{\partial K}\bigg|_{K=x}. \tag{3.46}$$

and the tails can be estimated using the same arguments from (3.46) – left tail by estimate of $F_{Q_{t,S_T}}(K_{[min]})$ and the right tail by 1- $F_{Q_{t,S_T}}(K_{[max]})$ where $K_{[min]}$ is the minimal and $K_{[max]}$ the maximal obseved strikes. Note that (3.46) can be also expressed in terms of IV and its derivatives, see (B.12).

In the next section we comment the application of these ideas to the two dimensional smoothing – recovering of the whole IV Surface – function of strike (or future moneyness) and time-to-maturity.

3.8.2 Estimating the IV-Surface

In the previous section we have considered the estimation of the IV-function for a single maturity. The aim of this section is to develop a technique for (arbitrage-free) estimation for any maturity, i.e. for the 2-dimentional IV-surface $\sigma(K,\tau)$.

The condition on the non-negative SPD can be taken from (3.43). The arbitrage in 'τ direction' is often referred to as calendar arbitrage. As already discussed in the section 2.4.2, assuming zero interest-rate, the so called total-variance $w(K,\tau) \overset{\text{def}}{=} \sigma^2(K,\tau)\tau$ is strictly increasing in τ. Assuming a deterministic time-varying interest rate, this translates to the argument that $w(\kappa,\tau) \overset{\text{def}}{=} \sigma^2(\kappa,\tau)\tau$ should be strictly increasing in τ under no-arbitrage. Assuming the model (3.37):

$$\tilde{\sigma}_i = \sigma(K_i,\tau_i) + \varepsilon_i \ , i = 1,\ldots,T, \tag{3.47}$$

our aim is to use the condition on the total variance in our smoothing estimate. Let us first introduce the two dimensional local polynomial estimator.

The idea of local polynomial estimation in higher dimensions is a straightforward generalization of the one-dimensional case. A standard – unconstrained two-dimensional local quadratic estimator $\hat{\sigma}(\kappa,\tau)$ is given by α_0

solving:

$$\min_{\alpha} \sum_{i=1}^{T} \mathcal{K}_H(\kappa-\kappa_i,\tau-\tau_i)\left\{\tilde{\sigma}_i-\alpha_0-\alpha_1(\kappa_i-\kappa)\right. \tag{3.48}$$
$$\left.-\alpha_2(\tau_i-\tau)-\alpha_{1,1}(\kappa_i-\kappa)^2-\alpha_{1,2}(\kappa_i-\kappa)(\tau_i-\tau)-\alpha_{2,2}(\tau_i-\tau)^2\right\}^2$$

where $\mathcal{K}_H(u) \stackrel{\text{def}}{=} \frac{1}{\det H}\mathcal{K}(H^{-1}u)$ is a (bivariate) kernel function with bandwidth-matrix H. Comparing (3.49) with a truncated bi-variate Taylor expansion of $\sigma(\kappa,\tau)$ shows $\alpha_0=\hat{\sigma}(\kappa,\tau)$, $\alpha_1=\frac{\partial\hat{\sigma}}{\partial\kappa}(\kappa,\tau)$, $\alpha_2=\frac{\partial\hat{\sigma}}{\partial\tau}(\kappa,\tau)$, $\alpha_{1,1}=\frac{\partial^2\hat{\sigma}}{2\partial\kappa^2}(\kappa,\tau)$, $\alpha_{2,2}=\frac{\partial^2\hat{\sigma}}{2\partial\tau^2}(\kappa,\tau)$, $\alpha_{1,2}=\frac{\partial^2\hat{\sigma}}{\partial\kappa\partial\tau}(\kappa,\tau)$. Since in our application it is typical to have small number of design points in the τ direction, we propose a parsimonious smoother $\hat{\sigma}(\kappa,\tau)$ given by α_0 – the solution of:

$$\min_{\alpha} \sum_{i=1}^{T} \mathcal{K}_H(\kappa-\kappa_i,\tau-\tau_i)\left\{\tilde{\sigma}_i-\alpha_0-\alpha_1(\kappa_i-\kappa)\right. \tag{3.49}$$
$$\left.-\alpha_2(\tau_i-\tau)-\alpha_{1,1}(\kappa_i-\kappa)^2-\alpha_{1,2}(\kappa_i-\kappa)(\tau_i-\tau)\right\}^2.$$

The idea of (3.49) is to construct a local smoother quadratic in κ and linear in τ.

Again the unconstrained estimate (3.49) may yield an estimate that contradicts the no-arbitrage assumptions. Our aim is to solve (3.49) w.r.t. non negative corresponding SPD and total variance strictly increasing in τ.

Consider first a problem of estimating the IV function for fixed τ which is not observed in the data set. Define $\hat{w}(\kappa,\tau)=\hat{\sigma}^2(\kappa,\tau)\tau$. Since $\frac{\partial\hat{w}}{\partial\tau}>0$ can be rewritten as $2\tau\alpha_0\alpha_2+\alpha_0^2>0$ for a given (single) τ we need solve the optimization problem (3.49) constrained by:

$$\begin{aligned}\hat{q}(\kappa,\tau)=\sqrt{\tau}\phi(d_1)\left\{\frac{1}{\kappa^2\alpha_0\tau}+\frac{2d_1}{\kappa\alpha_0\sqrt{\tau}}\alpha_1+\frac{d_1d_2}{\alpha_0}\alpha_1^2+2\alpha_{1,1}\right\} &\geq 0\\ 2\tau\alpha_0\alpha_2+\alpha_0^2 &> 0\end{aligned} \tag{3.50}$$

where $d_1=\frac{\alpha_0^2\tau/2-\ln(\kappa)}{\alpha_0\sqrt{\tau}}$, $d_2=d_1-\alpha_0\sqrt{\tau}$ and $F_\tau=S.e^{r\tau}$, for given (but arbitrary) κ.

If we are interested in estimating the entire IV-surface $\hat{\sigma}(\kappa,\tau)$ for a set of maturities $\{\tau_1,\ldots,\tau_L\}$ and for given value κ, we need to ensure $\hat{w}(\kappa,\tau_l,)\leq$

$\hat{w}(\kappa, \tau_{l'})$, for all $\tau_l < \tau_l'$. This leads to the following optimization problem:

$$\begin{aligned}
&\min_{\alpha(l)} \sum_{l=1}^{L} \sum_{i=1}^{T} \mathcal{K}_H(\kappa - \kappa_i, \tau_l - \tau_i) \{\tilde{\sigma}_i - \alpha_0(l) \\
&-\alpha_1(l)(\kappa_i - \kappa) - \alpha_2(l)(\tau_i - \tau) - \alpha_{1,1}(l)(\kappa_i - \kappa)^2 \\
&- \alpha_{1,2}(l)(\kappa_i - \kappa)(\tau_i - \tau)\}^2
\end{aligned} \tag{3.51}$$

subject to

$$\begin{aligned}
\sqrt{\tau_l}\phi(d_1(l)) \left\{ \frac{1}{\kappa^2 \alpha_0(l) \tau_l} + \frac{2 d_1(l)}{\kappa \alpha_0(l) \sqrt{\tau_l}} \alpha_1(l) + \frac{d_1(l) d_2(l)}{a_0(l)} \alpha_1^2(l) + 2\alpha_{1,1}(l) \right\} &\geq 0, \\
d_1(l) = \frac{\alpha_0^2(l)\tau_l/2 - \ln(\kappa)}{\alpha_0(l)\sqrt{\tau_l}}, \quad d_2(l) = d_1(l) - a_0(l)\sqrt{\tau_l}, \quad l = 1, \ \ldots \ , L & \\
2\tau_l \alpha_0(l) \alpha_2(l) + \alpha_0^2(l) > 0 \quad l = 1, \ \ldots \ , L & \\
\alpha_0^2(l)\tau_l < \alpha_0^2(l')\tau_l', \quad \tau_l < \tau_l'. &
\end{aligned} \tag{3.52}$$

Comparing (3.51) – (3.52) with the one-dimensional problem, (3.51) – (3.52) calculates for given κ the estimates for all given τ_l in one step in order to guarantee increasing $\hat{w}$ in τ. The bivariate kernel function $\mathcal{K}_H(\kappa - \kappa_i, \tau_l - \tau_i)$ is given by the product of two univariate kernel functions: $\mathcal{K}_{h_\kappa}(\kappa - \kappa_i) = \frac{1}{h_\kappa}\mathcal{K}\left(\frac{\kappa - \kappa_i}{h_\kappa}\right)$, $\mathcal{K}_{h_\tau}(\tau - \tau_i) \stackrel{\text{def}}{=} \frac{1}{h_\tau}\mathcal{K}\left(\frac{\tau - \tau_i}{h_\tau}\right)$ where $\mathcal{K}$ is the Epanechnikov kernel.

Figure 3.11 shows the results for the daily data on February 2, 2006. By analogy to the univariate case, we consider global $h_\kappa = 0.05$. The h_τ is chosen increasing as τ increases $h_\tau = 0.2$ for $0 < \tau \leq \frac{1}{3}$, 0.3 for $\frac{1}{3} < \tau \leq \frac{2}{3}$ and 0.4 for $\frac{2}{3} < \tau \leq 1$. This choice was made again in such a way to obtain sufficient number of data for estimating parameters. Problem (3.51) - (3.52) was solved in system GAMS 22.0 - solver MINOS, for each κ separately. Similar to one-dimensional problem, a more sophisticated choice of the smoothing parameter H can be done by considering the cross-validation principle as mentioned in the one-dimensional case, however, since in general situation, we need to optimize cross-validation criterion w.r.t. to 2×2 matrix, this choice of smoothing parameters is computationally much more involved.

As an alternative to estimating the IV surface by smoothing the IVs, we may smooth in total variance w directly. This can be done via the same route, since the SPD can be expressed in terms of the total variance w. From the definition of w follows: $d_1 = \sqrt{w}/2 - \ln(\kappa)/\sqrt{w}$, $d_2 = -\ln(\kappa)/\sqrt{w} - \sqrt{w}/2$ and

$$q(\kappa, \tau) = \frac{\sqrt{\tau}\phi(d_1)}{\sqrt{w\tau}} \left\{ \frac{1}{\kappa^2} + \frac{d_1}{k\sqrt{w}} \left(\frac{\partial w}{\partial k}\right) + \frac{d_1 d_2 - 1}{4w} \left(\frac{\partial w}{\partial k}\right)^2 + \frac{1}{2} \left(\frac{\partial^2 w}{\partial k^2}\right) \right\}$$

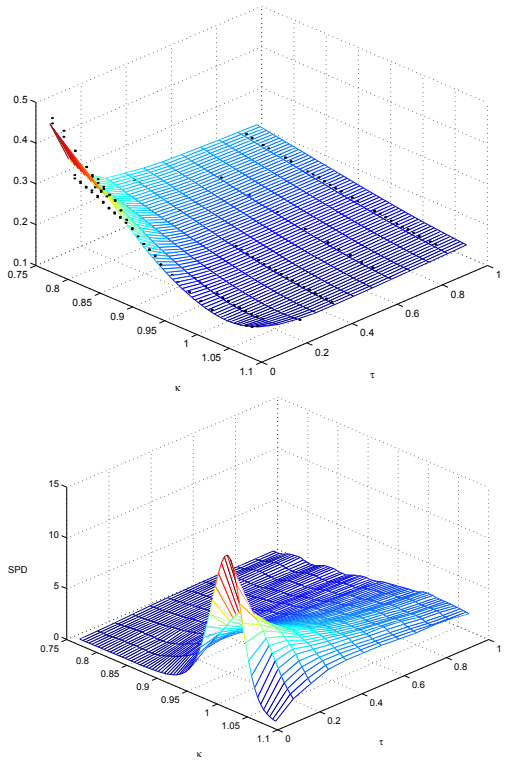

Figure 3.11: Left figure: Smoothed IV surface – left figure and corresponding family of SPDs, daily data, February 2, 2006, horizontal axis are moneyness (κ) and time-to-maturity (τ).

since $\partial w/\partial\kappa = 2\sigma\tau\partial\sigma/\partial\kappa$ and $\partial^2 w/\partial\kappa^2 = \frac{1}{2w}(\frac{\partial w}{\partial\kappa})^2 + 2\sqrt{w\tau}(\frac{\partial^2\sigma}{\partial\kappa^2})$ Using these expressions we can design an estimator for the IV-surface.

The 'observed' total variance is determined from the observed IVs by $\tilde{w}_i \stackrel{\text{def}}{=} \tilde{\sigma}_i^2\tau_i$. Assuming the observation model:

$$\tilde{w}_i = w(\kappa_i, \tau_i) + \varepsilon_i \tag{3.53}$$

$i = 1, \ldots n$, the estimate of the true total variance function $w(\kappa, \tau) \stackrel{\text{def}}{=} \sigma^2(\kappa, \tau)\tau$ at the point (κ, τ), denoted by $\hat{w}(\kappa, \tau)$, is given by solution of the following optimization problem:

$$\begin{aligned}
&\min_{\alpha} \sum_{l=1}^{L}\sum_{i=1}^{T} \mathcal{K}_H(\kappa - \kappa_i, \tau_l - \tau_j)\,\{\tilde{w}_i - \alpha_0(l) \\
&-\alpha_1(l)(\kappa_i - \kappa) - \alpha_2(l)(\tau_i - \tau_l) - \alpha_{1,1}(l)(\kappa_i - \kappa)^2 \\
&-\alpha_{1,2}(l)(\kappa_i - \kappa)(\tau_i - \tau)\}^2
\end{aligned} \tag{3.54}$$

subject to

$$\begin{aligned}
\frac{\sqrt{\tau_l}\phi(d_1(l))}{\sqrt{\alpha_0(l)\tau_l}}\left\{\frac{1}{\kappa^2} + \frac{d_1(l)}{\kappa\sqrt{\alpha_0(l)}}\alpha_1(l) + \frac{d_1(l)d_2(l) - 1}{4\alpha_0(l)}\alpha_1^2(l) + (\alpha_{1,1}(l))\right\} &\geq 0, \\
l = 1, &\ \ldots\ , L \\
d_1(l) = \sqrt{\alpha_0(l)}/2 - \ln(\kappa)/\sqrt{\alpha_0(l)}, &\quad l = 1, \ \ldots\ , L \\
d_2(l) = -\ln(\kappa)/\sqrt{\alpha_0(l)} - \sqrt{\alpha_0(l)}/2, &\quad l = 1, \ \ldots\ , L \\
\alpha_2(l) > 0, &\quad l = 1, \ \ldots\ , L \\
\alpha_0(l) < \alpha_0(l'),\ \tau_l &< \tau_l'
\end{aligned} \tag{3.55}$$

As we can see, considering the smoothing in the total variance the last two conditions $\alpha_0(l) < \alpha_0(l')$, $\tau_l < \tau_l'$ and $\alpha_2(l) > 0$, $l = 1, \ldots, L$ are simpler than the analogue condition expressed in terms of IV. In the last step we estimate the IV surface by setting $\hat{\sigma}(\kappa, \tau_l) \stackrel{\text{def}}{=} \sqrt{\hat{w}(\kappa, \tau_l)\tau_l^{-1}}$.

The comparative advantage to smoothing in IV is the simpler structure of the constrains. On the other hand, in this case, the IV surface must be calculated by setting $\hat{\sigma}(\kappa, \tau_l) = \sqrt{\hat{w}(\kappa, \tau_l)\tau_l^{-1}}$.

3.9 Further Comments on IV-smoothing

In this chapter several topics of the static estimation of the IV functions and surfaces have been discussed, mainly focused on the nonparametric (local

polynomial) estimation. However two topics, that have been omitted here, seems to the author worth to mention as possible further research. Firstly, one should keep in mind that IV is a transformation of the option price (strongly non-linear). One of the consequences is that the sensitivity of the option price to the BS-IV is not constant. The sensitivity measured in the partial derivative of the option price w.r.t. BS-IV is called BS-Vega (defined in the section 2.1) and plotted in the figure 3.12 for different $\tau \in [0.05, 1]$ and stock price $S \in [50, 150]$ and strike price $K = 100$, and is decreasing in τ. Figure 3.12 is borrowed from Fengler [2005b], where similar pictures for other Greeks can be found. Consequently the small change (error) in option price may lead to big change (error) in the BS-IV. One way-out can

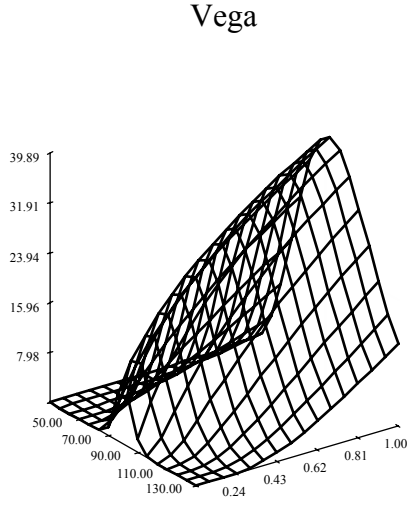

Figure 3.12: Example of vega of an option as a function of asset price and time-to-maturity for a fixed strike price $K = 100$.

be in-fact an estimate with additional weighting scheme based on the corresponding Vega. In fact Brunner and Hafner [2003] proposed a weighting scheme based on the "call-equivalent" ratio between vega and delta corresponding to the observed option in order to correct heteroscedasticity in a parametric approach of estimating IVs. This approach may be also used for correcting the "vega" problems in the IV estimation. Also robust estimation techniques may improve the performance and stability of the estimates.

Another challenge is the degenerated design in the τ direction – decreasing number of traded options in τ and increasing sparsiness of the design (for the first three months 3 possible time-to-maturities are available, afterwards in semiannual, and annual cycles). Of course the τ design can be transformed

by considering proper transformation, e.g. log-scale of days to maturity but this does not solve the problem completely. Promising proposal is the combination of the smoothing techniques with the dynamical factor models proposed by Fengler et al. [2005] and will be briefly discussed in the chapter 4. However, there is no study known to the author that would directly consider the special structure of the time-to-maturity design by estimating the IVS. Unfortunately, as it is visible from the section 3.8.2, even the financial theory does not provide so much information about the "term structure" in compare to the strike direction. The deeper theoretical results might be used by improving the performance of the estimates by ether constrained estimation presented above or by constructing semiparametric alternatives.

Chapter 4

Functional Data Analysis and Empirical Finance

In many different fields of applied statistics the object of interest is depending on some continuous parameter, e.g. continuous time. The typical examples in finance are, already discussed in previous chapters, the IV functions or yield curves. Due to different market conventions or other technical reasons these data are observable only on a discrete grid, e.g. on a grid of strikes and maturities for which the trade has been settled at a given time-point. However it is clear that IV can be calculated for any (positive) call-price and, since it is a monotonous mapping, IV is also a function of K and τ, and these parameters can be in general treated as continuous parameters. By collecting these functions for several time points (e.g. days) or for different underlyings a bunch (sample) of functions is obtained – *functional data set*. The questions arising by the statistical analysis of functional data are basically identical to the standard statistical analysis of univariate or multivariate objects. From a theoretical point of view, (stochastic) models and statistical analysis of a functional data set – *Functional Data Analysis* can be taken often one-to-one from the conventional multivariate analysis. In fact the first method how to deal with the functional data is to discretize them and perform a standard multivariate analysis on the resulting random vectors.

This short introductions yields a natural question: "what makes a data set functional" – when is it appropriate to understand a vector $v_j, j = 1, \ldots, d$ as a vector of discretized functional values $v(x_j) = v_j,\ j = 1, \ldots, d$? Heuristically, the answer can be very simple: then, when there is a reason to assume that the function v possess some features that are meaningful only in the functional context. Probably the most important example of such a feature is the *smoothness*. Smoothness has no direct meaning for vectors but is well defined for functions. Next important example that needs to be considered

as a functional data, even if the data are strongly discretized, are samples of probability density functions (e.g. SPD in the finance).

This chapter is organized as follows: Section 4.1 briefly introduces the basic setup of the functional data analysis, section 4.2 defines the functional counterpart of the well known principal components method motivated by the Karhunen-Loève decomposition – functional principal components. The issues associated with the implementation, especially using the basis expansion approach, and regularized principal components proposed by Silverman [1996] are discussed. The methods based on the functional basis expansion together with some examples on visualization of the FDA objects are applied to the descriptive analysis of EURIBOR yield curves in the section 4.2.5. An approach of estimating the functional principal components motivated by the duality relations between column and row spaces of sample matrices known from multivariate PCA, proposed by Benko et al. [2006b] is presented in the section 4.2.6. Connection between two sample problem and FPCA for two different samples is studied in the section 4.3 and the application of these methods on the study of the term structure of IV surfaces is given in the section 4.5. The implementation of the library for handling the functional data analysis for statistical computing environment XploRe is presented and discussed in Appendix A. This chapter follows partially the arguments presented in Benko et al. [2006b], Benko and Härdle [2005], Benko [2004] and Ramsay and Silverman [2005].

4.1 Basic Setup of Functional Data Analysis

This section introduces basic technical tools for analysis of the random functions, discuss the step from multivariate to functional data and shows the parallels. A certain knowledge of metric spaces and the theory of measure and integral is expected (e.g. the vector space, completeness or Borel sets are not explicitly defined here, see Lukeš and Malý [1995] and Dupačová et al. [2002] for reference among others).

In the traditional multivariate framework a random object is modeled through a d-dimensional random vector X – is modeled as a measurable function

$$X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}^d),$$

mapping a probability space $(\Omega, \mathcal{A}, P)$ to the real measurable space $(\mathbb{R}^d, \mathcal{B}^d)$, where $\mathcal{B}^d$ are the Borel sets on $\mathbb{R}^d$.

An d-dimensional real space $\mathbb{R}^d$ together with standard scalar product $\langle x_l, x_k \rangle = x_l^\top x_t$ for $x_l, x_k \in \mathbb{R}^d$ is an example of *Hilbert Space*. Formally vector

space H equipped with a scalar product $\langle .,. \rangle$ is called Hilbert space if it is complete w.r.t. norm generated by the scalar product $||u|| = \sqrt{\langle u, u \rangle}$, $u \in H$.

Using this notation, a random object in FDA can be modeled as a measurable function

$$X : (\Omega, \mathcal{A}, P) \to (H, \mathcal{B}_H),$$

where H is some (separable) Hilbert space of functions and $\mathcal{B}_H$ stands for Borel field.

For the purpose of FDA, the most often used function space is the space of the Lebesque integrable functions - L_J^2 on some subset of $J \subseteq \mathbb{R}^q$, i.e. the space of real-valued functions with support in J such that $(\int_J f^2(u)du)^{1/2} < \infty$, $\langle w, v \rangle_{L^2} = \int_J w(u)v(u)du$. The norm generated by the scalar product $\langle w, v \rangle_{L^2}$ is denoted by $||u||_{L^2} = \sqrt{\langle u, u \rangle_{L^2}}$.

In the following text the $\langle u, v \rangle$ and $||u||$ denotes the standard Euclidean scalar product and norm for $u \in \mathbb{R}^d$ and L_2 norm for functions in order to simplify the notation.

Note that the FDA deals typically with smooth functions, this would formally lead to more complicated type of function spaces – *Sobolev spaces* – spaces of smooth Lebesque integrable functions, however with couple of exceptions the results on Sobolev spaces are not explicitly used in this chapter and hence are not discussed in detail at this place, for a recent work on empirical finance using Sobolev spaces, see Pešta [2006].

Assume the existence of the expected value, variance and continuous covariance and correlation function of (functional) random variable X, and denote these by $\mu(t)$, $\varsigma(t)$, $\sigma(s,t)$ and $\rho(s,t)$ respectively:

$$\begin{aligned}
\mu(t) &= EX(t),\ t \in J, \\
\varsigma(t) &= E\{X(t) - \mu(t)\}^2,\ t \in J, \\
\sigma(s,t) &= E\{X(s) - \mu(s)\}\{X(t) - \mu(t)\},\ s,t \in J, \\
\rho(s,t) &= \frac{\sigma(s,t)}{\sqrt{\varsigma(s)\varsigma(t)}}\ s,t \in J.
\end{aligned}$$

The $\rho(s,t)$ is defined under the assumption $\varsigma(s), \varsigma(t) > 0$. Then $E(\|X - \mu\|^2) = \int \sigma(t,t)dt < \infty$, and the covariance operator Γ of X is given by

$$(\Gamma v)(t) = \int \sigma(t,s)v(s)ds, \quad v \in L_J^2. \tag{4.1}$$

Note that in order to distinguish between covariance function of two functional variables and covariance function, $\sigma(s,t)$, as defined above, $\sigma(s,t)$ is sometimes called cross-covariance function. Since we do not consider two

different random variables in this chapter, we will use the shorter version in the following text.

For the functional sample $X_i(t),\ i = 1, \dots n$ an estimates of $\mu(t)$, $\varsigma(t)$, $\sigma(s,t)$ and $\rho(s,t)$ are constructed as straightforward generalizations of the multivariate counterparts:

$$\begin{aligned}
\bar{X}(t) &= \tfrac{1}{n}\sum_{i=1}^{n} X_i(t), \\
\widehat{\varsigma}(t) &= \tfrac{1}{n}\sum_{i=1}^{n} \left\{X_i(t) - \bar{X}(t)\right\}^2, \\
\widehat{\sigma}(s,t) &= \tfrac{1}{n}\sum_{i=1}^{n} \left\{X_i(s) - \bar{X}(s)\right\}\left\{X_i(t) - \bar{X}(t)\right\}, \\
\widehat{\rho}(s,t) &= \tfrac{\widehat{\sigma}(s,t)}{\sqrt{\widehat{\varsigma}(s)\widehat{\varsigma}(t)}}.
\end{aligned}$$

The point-wise consistency of these estimators can be obtained using standard multivariate results. The covariance operator (4.1) can be approximated by the empirical covariance operator

$$(\hat{\Gamma}_n v)(t) = \int \hat{\sigma}(t,s)v(s)ds. \tag{4.2}$$

In fact all multivariate techniques can be transferred to the FDA: the basic descriptive statistics described above, regression models (linear, generalized,...), canonical analysis, see monographs Ramsay and Silverman [2005] and Ramsay and Silverman [2002] up to the nonparametric methods recently summarized by Ferraty and Vieu [2006]. The main focus here is the Principal Component Analysis in the FDA. The functional version of principal component analysis has much more important role than its multivariate version, in fact it is often only way to describe and work with distribution of random functions in practice.

4.2 Principal Components for Functional Data

Principal Components Analysis (PCA) yields dimension reduction in the multivariate framework. The aim is to find a normalized weight vectors $\gamma_r \in \mathbb{R}^d$ for which the linear transformations of a T-dimensional random vector $\mathbf{x}$:

$$\beta_r = \gamma_r^\top(\mathbf{x} - E\mathbf{x}) = \ \langle \gamma_r, \mathbf{x} - E\mathbf{x}\rangle, \tag{4.3}$$

have maximal variance subject to:

$$\gamma_l^\top \gamma_r = \ \langle \gamma_l, \gamma_r \rangle \ = \boldsymbol{I}(l = r) \text{ for } l \leq r.$$

The problem is solved by the means of the Jordan spectral decomposition of the covariance matrix, Härdle and Simar [2003], page 63 – the r-th principal component is the eigenvector of covariance matrix corresponding to the r-th largest eingenvalue.

Denoting the d-dimensional sample of range n by $\mathcal{X}$ and assuming without loss of generality that the sample is centered, the eigenvectors and eigenvalues can be estimated by the eigenvalues and eigenvectors of the sample covariance matrix

$$\mathcal{C} = \frac{1}{n}\mathcal{X}^\top\mathcal{X}.$$

It can be easily seen that there is a simple connection between the column and row space of the matrix $\mathcal{X}$. Denoting the eigenvectors of $\mathcal{X}^\top\mathcal{X}$ by u_k and eigenvectors of $\mathcal{X}\mathcal{X}^\top$ by v_k, for $k = 1, \ldots, R = \text{rank}(\mathcal{X})$, following duality relations holds:

$$u_k = \frac{1}{\sqrt{l_k}}\mathcal{X}^\top v_k, \tag{4.4}$$

$$v_k = \frac{1}{\sqrt{l_k}}\mathcal{X}u_k, \tag{4.5}$$

where l_k is the $k-th$ largest eigenvalue of $\mathcal{X}^\top\mathcal{X}$. Moreover the non-zero eingenvalues of $\mathcal{X}^\top\mathcal{X}$ and $\mathcal{X}\mathcal{X}^\top$ are equal.

In the FDA the motivation for Functional Principal Components Analysis (FPCA) as the dimension reduction technique can be done via the same route: having a random function X, find orthonormal weight functions $\gamma_1, \gamma_2, \ldots$, such that the variance of the linear transformation

$$\beta_r = \langle \gamma_r, X - \mu \rangle = \int \gamma_r(t)\{X(t) - \mu(t)\}dt, \tag{4.6}$$

is maximal, w.r.t. to the orthonormal weight functions γ_r, i.e. such that:

$$||\gamma_r||^2 = \int \gamma_r(t)^2 dt = 1,$$

$$\langle \gamma_l, \gamma_r \rangle = \int \gamma_l(t)\gamma_r(t)dt = 0, \; l \neq r.$$

Summarizing, the desired weight functions solve:

$$\underset{\langle \gamma_l, \gamma_r \rangle = \boldsymbol{I}(l=r), l \leq r}{\arg\max} Var\langle \gamma_r, X - \mu \rangle, \tag{4.7}$$

or equivalently:

$$\underset{\langle \gamma_l, \gamma_r \rangle = \boldsymbol{I}(l=r), l \leq r}{\arg\max} \int\int \gamma_r(s)\sigma(s,t)\gamma_r(t)dsdt.$$

The solution of the optimization criterion (4.7) is obtained by solving the Fredholm functional eigenequation

$$\int \sigma(s,t)\gamma(t)dt = \lambda\gamma(s). \tag{4.8}$$

The solution is achieved by γ_r – eigenfunction of the covariance operator Γ in (4.1) corresponding to the r-th largest eigenvalue λ_r.

4.2.1 Karhunen-Loève Expansion

The theoretical basis for the application of the FPCA motivated above as the dimension reduction tool is given by Karhunen-Loève Expansion (KL). KL expansion of a random function X is obtained by:

$$X = \mu + \sum_{r=1}^{\infty} \beta_r \gamma_r, \tag{4.9}$$

The factor loadings $\beta_r = \langle X - \mu, \gamma_r \rangle$ are uncorrelated (scalar) factor loadings with $E(\beta_r) = 0$, $E(\beta_r^2) = \lambda_r$, and $E(\beta_r \beta_k) = 0$ for $r \neq k$.

Clearly, the KL expansion provides a basic tool to access the distribution of a (random) function. The structure of the distribution of X can be analyzed by analyzing the structure of the eigenfunctions (γ_r) and of the (one-dimensional) factor loading β_r.

The Properties of KL-Expansion are following:

- Convergence: Under the assumptions given here the sum in KL-Expansion converges with probability 1.

- Smoothness: If the X is a function with realizations that are twice continuously differentiable with probability 1 then this will be also true for the eigenfunctions γ_r.

- Best Empirical Basis: First L principal components provide a "best basis" for approximating the sample functions in terms of the integrated square error. For any choice of L orthonormal basis functions $v_1, \ldots, v_L$ the mean integrated square error:

$$\varrho(v_1, \ldots, v_L) = E(\| X - \mu - \sum_{r=1}^{L} \langle X - \mu, v_r \rangle v_r \|^2) \tag{4.10}$$

  is minimized by $v_r = \gamma_r$.

Especially the "Best Empirical Basis" property (4.10) is of high importance in the application of the FPCA. In many important applications a small number of functional principal components will suffice to approximate random function X with a small (residual) error (4.10). In the case of Gaussian random function X the β_r are also Normally distributed. Clearly this implies $\beta_r \sim N(0, \lambda_r)$ moreover β_r, β_k are independent for $r \neq k$. In this case, assuming that truncated KL-expansion:

$$\tilde{X} = \mu + \sum_{r=1}^{L} \beta_r \gamma_r, \tag{4.11}$$

approximates X sufficiently well, (4.11) can serve as a simply implementable simulation tool of the random functions X – knowing the eigenfunctions γ_r only the one-dimensional normally distributed variables $\beta_r \sim N(0, \lambda_r),\ r = 1, \ldots, L$ need to be simulated. Clearly simulating the scalar random variables is much more simpler than simulation of random functions.

4.2.2 Estimation of Functional Principal Components

The estimation of FPC is done using same arguments as in multivariate PCA: for a given sample X_i of range n generated by X an empirical analog of (4.9) can be constructed by using eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots$ and orthonormal eigenfunctions $\hat{\gamma}_1, \hat{\gamma}_2, \ldots$ of the empirical covariance operator $\hat{\Gamma}_n$, see 4.2. Then

$$X_i = \bar{X} + \sum_{r=1}^{n} \hat{\beta}_{ri} \hat{\gamma}_r, \quad i = 1, \ldots, n, \tag{4.12}$$

where $\hat{\beta}_{ri} =< \hat{\gamma}_r, X_i - \bar{X} >$. Furthermore $n^{-1} \sum_i \hat{\beta}_{ri} = 0$, $n^{-1} \sum_i \hat{\beta}_{ri} \hat{\beta}_{si} = 0$ for $r \neq s$, and $n^{-1} \sum_i \hat{\beta}_{ri}^2 = \hat{\lambda}_r$. Obviously, $\hat{\lambda}_r$ and $\hat{\gamma}_r$ estimate λ_r and γ_r for $r = 1, 2, \ldots$. The asymptotic behavior of these estimates is stated in the following theorem.

THEOREM 4.1 *Assume that $X_1, \ldots, X_n \in L^2[0,1]$ are an i.i.d. sample of random functions with mean μ and continuous covariance function $\sigma(t,s)$, and (4.9) holds for a system of eigenfunctions satisfying $\sup_r \sup_{t\in[0,1]} \gamma_r(t) < \infty$.*

Furthermore, $\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} E[\beta_{ri}^2 \beta_{si}^2] < \infty$ and $\sum_{q=1}^{\infty} \sum_{s=1}^{\infty} E[\beta_{ri}^2 \beta_{qi} \beta_{si}] < \infty$ for all $r = 1, 2, \ldots$.

i) For all $t \in [0,1]$

$$\sqrt{n}\{\bar{X}(t) - \mu(t)\} = \sum_r \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \beta_{ri} \right\} \gamma_r(t) \xrightarrow{\mathcal{L}} N\left(0,\ \sum_r \lambda_r \gamma_r(t)^2\right),$$

ii) If, furthermore, $\lambda_{r-1} > \lambda_r > \lambda_{r+1}$ *holds for some fixed* $r \in \{1, 2, \dots\}$, *then*

$$\sqrt{n}(\hat{\lambda}_r - \lambda_r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\beta_{ri}^2 - \lambda_r\}\right] + \mathcal{O}_p(n^{-1/2}) \xrightarrow{\mathcal{L}} N(0, \Lambda_r), \quad (4.13)$$

where $\Lambda_r = E[(\beta_{ri}^2 - \lambda_r)^2]$,

iii) and for all $t \in [0, 1]$

$$\hat{\gamma}_r(t) - \gamma_r(t) = \sum_{s \neq r} \left\{ \frac{1}{n(\lambda_r - \lambda_s)} \sum_{i=1}^{n} \beta_{si}\beta_{ri} \right\} \gamma_s(t) + R_r(t), \quad (4.14)$$

where $\|R_r\| = \mathcal{O}_p(n^{-1})$.

Moreover, $\sqrt{n} \sum_{s \neq r} \left\{ \frac{1}{n(\lambda_r - \lambda_s)} \sum_{i=1}^{n} \beta_{si}\beta_{ri} \right\} \gamma_s(t) \xrightarrow{\mathcal{L}}$
$N\left(0,\ \sum_{q \neq r} \sum_{s \neq r} \frac{E[\beta_{ri}^2 \beta_{qi} \beta_{si}]}{(\lambda_q - \lambda_r)(\lambda_s - \lambda_r)} \gamma_q(t)\gamma_s(t)\right).$

Proof:
The proof can be found in Benko et al. [2006b]. □

COROLLARY 4.1 *Assume that* $X_1, \dots, X_n \in L^2[0, 1]$ *are an i.i.d. sample of Gaussian random functions with mean* μ *and continuous covariance function* $\sigma(t, s)$, *and (4.9) holds for a system of eigenfunctions satisfying* $\sup_r \sup_{t \in [0,1]} \gamma_r(t) < \infty$. *Furthermore,* $\lambda_{r-1} > \lambda_r > \lambda_{r+1}$ *holds for some fixed* $r \in \{1, 2, \dots\}$, *then*

i) $\sqrt{n}(\hat{\lambda}_r - \lambda_r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\beta_{ri}^2 - \lambda_r\}] + \mathcal{O}_p(n^{-1/2}) \xrightarrow{\mathcal{L}} N(0, 2\lambda_r^2)$,

ii) Moreover, $\sum_{q \neq r} \sum_{s \neq r} \frac{E[\beta_{ri}^2 \beta_{qi} \beta_{si}]}{(\lambda_q - \lambda_r)(\lambda_s - \lambda_r)} \gamma_q(t)\gamma_s(t) = \sum_{s \neq r} \frac{\lambda_r \lambda_s}{(\lambda_s - \lambda_r)^2} \gamma_s(t)^2$.

Proof:
In case of Gaussian random functions the factor loadings are also Gaussian. Then β_{ri} and β_{si} are independent for $r \neq s$, all moments of moments β_{ri} are finite, and hence $E[\beta_{ri}^2 \beta_{qi} \beta_{si}] = 0$ for $q \neq s$ as well as $E[\beta_{ri}^2 \beta_{si}^2] = \lambda_r \lambda_s$ for $r \neq s$, see Gihman and Skorohod [1973]. The explicit asymptotic results for Gaussian case were first derived by Dauxois et al. [1982]. □

COROLLARY 4.2 *Assume that the assumptions of the theorem 4.1 are fulfilled. Then*

$$\begin{aligned} \| \hat{\gamma}_r - \gamma_r \| &= \mathcal{O}_p(n^{-1/2}), \\ |\hat{\lambda}_r - \lambda_r| &= \mathcal{O}_p(n^{-1/2}), \\ |\hat{\beta}_{ri} - \beta_{ri}| &= \mathcal{O}_p(n^{-1/2}). \end{aligned}$$

Proof:
Direct consequence of theorem 4.1. These results corresponds to the results derived by Dauxois et al. [1982]. □

4.2.3 Implementation via Basis Expansion

A popular way of FDA-implementation is to use a truncated functional basis expansion. More precisely, denote a functional basis on the interval J by $\{\theta_1, \theta_2, \dots, \}$ and assume that the functions X_i are approximated by the first L basis functions θ_l, $l = 1, 2, \dots L$, i.e. by linear combination of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_L)^\top$:

$$X_{\theta;i}(t) = \sum_{l=1}^{L} c_{il}\theta_l(t) = \mathbf{c}_i^\top \boldsymbol{\theta}(t), \tag{4.15}$$

where $\mathbf{c}_i = (c_{i1}, \dots, c_{iL})^\top$. Note that this section is in fact recapitulation of the section 3.3 and clarification of the notation in the FDA.

Using the functional basis expansion, the analysis of the functional objects is implemented through the coefficient matrix

$$\mathbf{C} = \{c_{il},\ i = 1, \dots, n,\ l = 1, \dots, L\},$$

e.g. the mean, variance, covariance and correlation functions can be approximated by:

$$\begin{aligned} \bar{X}_\theta(t) &= \bar{\mathbf{c}}^\top \boldsymbol{\theta}(t), \\ \widehat{\varsigma}_\theta(t) &= \boldsymbol{\theta}(t)^\top Cov(\mathbf{C})\boldsymbol{\theta}(t), \\ \widehat{\sigma}_\theta(s,t) &= \boldsymbol{\theta}(s)^\top Cov(\mathbf{C})\boldsymbol{\theta}(t), \\ \widehat{\rho}_\theta(s,t) &= \frac{\widehat{\sigma}(s,t)}{\{\widehat{\varsigma}(t)\widehat{\varsigma}(s)\}^{1/2}} \end{aligned}$$

where $\bar{\mathbf{c}}$ is a vector with following elements:

$$\bar{c}_l = \frac{1}{n}\sum_{i=1}^{n} c_{il},\ l = 1, \dots, L,$$

$$Cov(\mathbf{C}) = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{c}_i - \bar{\mathbf{c}})(\mathbf{c}_i - \bar{\mathbf{c}})^\top.$$

The scalar product of two functions $\langle X_i, X_j \rangle$ is approximated by:

$$\langle X_{\theta,i}, X_{\theta,j} \rangle \ = \int X_{\theta,i}(t) X_{\theta,j}(t) dt = \mathbf{c}_i^\top \mathbf{W} \mathbf{c}_j,$$

where

$$\mathbf{W} \stackrel{\text{def}}{=} \int \boldsymbol{\theta}(t)\boldsymbol{\theta}(t)^\top dt. \tag{4.16}$$

Approximation and Coefficient Estimation

The coefficient matrix $\mathbf{C}$ has to be estimated from the data:

$$Y_{ik} = X_i(t_{ik}) + \varepsilon_{ik} \tag{4.17}$$

for $k = 1, \ldots T_i$ and $i = 1, \ldots, n$ with $E(\varepsilon_{ik}) = 0$, $Var(\varepsilon_{ik}) = \sigma_i$. Clearly (4.17) corresponds to n regression problems (3.1), and elements of the coefficient vectors $\mathbf{c}_i$ for $i = 1, \ldots, n$ can be estimated by techniques introduced in section 3.3.4.

Note that strictly speaking in the practice two sources of randomness are present in the FPCA problems. First is the difference between eigenfunction of the true covariance operator – γ_r and the eigenfunctions of the sample covariance operator – $\hat{\gamma}_r$ and the difference between $\hat{\gamma}_r$ and the $\hat{\gamma}_{\theta,T,r}$ – eigenfunctions of the covariance operator based on the functions expanded (or approximated) by the basis $\boldsymbol{\theta}$.

Note that in the current section the functional basis approach by estimating the functions X_i and consequently the eigenfunctions γ_r is employed, any of the nonparametric estimation methods described in chapter 3 can be considered. An alternative approach based on the local regression techniques is considered in the section 4.2.6.

In the standard FDA setup, one assumes that the functions X_i are observed without additional error and the analysis proceeds "as if" the functions were directly observed. Obviously this assumption is often violated in practice and this approach is justified only if the additional noise is of smaller order in compare to the variation of the functions X_i. For the purposes of this section, mainly in order to avoid complicated notation, we will neglect this difference. However, from a statistical point of view we should keep this difference in mind, for more detailed discussion see section 4.2.6.

Note also, that using the functional basis approach, the functional problem (of infinity dimension) is "boiled down" to the finite dimensional vector

problem, the functional scalar product is approximated by a weighted scalar product of the coefficient vectors (4.16). This is also one of the reasons for the popularity of this approach – the FDA can be often handled by slightly adjusted multivariate procedures. A nice example is the implementation of the FPCA using functional basis expansion technique.

Implementation of Functional Principal Components

Suppose that the weight function γ is approximated sufficient well by γ_θ with the following expansion w.r.t. functional basis $\boldsymbol{\theta}$:

$$\gamma_\theta = \sum_{l=1}^{L} b_l \theta_l = \boldsymbol{\theta}^\top \mathbf{b},$$

where $\boldsymbol{b} = (b_1, b_2, \ldots, b_L)^\top$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)$.

Using this notation the left hand side of eigenequation (4.8) can be rewritten:

$$\begin{aligned} \int \widehat{\sigma}(s,t)\gamma(t)dt &\approx \int \boldsymbol{\theta}(s)^\top Cov(\mathbf{C})\boldsymbol{\theta}(t)\boldsymbol{\theta}(t)^\top \mathbf{b} dt \\ &= \boldsymbol{\theta}^\top Cov(\mathbf{C})\mathbf{W}\mathbf{b}, \end{aligned}$$

so that:

$$Cov(\mathbf{C})\mathbf{W}\mathbf{b} = \lambda \mathbf{b}.$$

The functional scalar product $\langle \gamma_l, \gamma_k \rangle$ corresponds to $\mathbf{b}_l^\top \mathbf{W} \mathbf{b}_k$. Matrix $\mathbf{W}$ is symmetric by definition, thus, defining $\mathbf{u} = \mathbf{W}^{1/2}\mathbf{b}$, one needs to solve finally a symmetric eigenvalue problem:

$$\mathbf{W}^{1/2} Cov(\mathbf{C}) \mathbf{W}^{1/2} \mathbf{u} = \lambda \mathbf{u},$$

and to compute the inverse transformation $\mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}$. For the orthonormal functional basis (e.g. for Fourier basis) $\mathbf{W} = \mathbf{I}$, i.e. the problem of FPCA is reduced to the multivariate PCA performed on the matrix $\mathbf{C}$.

Algorithm

1. calculate $\mathbf{C}$ and $\mathbf{W}$
2. using Cholesky decomposition calculate $\mathbf{W}^{1/2}$
3. use symmetric matrix eigenvalue routine and obtain eigenvalues and eigenvectors ($\mathbf{u}$) of $\mathbf{W}^{1/2} Cov(\mathbf{C}) \mathbf{W}^{1/2}$

4. calculate $\mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}$

Note that, the estimated coefficient of eigenfunctions are orthonormal with respect to the scalar product $\mathbf{b}_i^\top \mathbf{W}\mathbf{b}_j$, however due to some numerical errors there can by small deviances.

4.2.4 Smoothed Functional Principal Components (SFPCA)

The idea of combining an optimization criterion like explained variance with some type of roughness penalty in order to achieve a regularization or smoothness of the results can be used quite generally in the functional data analysis. It is also a very useful techniques in combination with FPCA.

Assume that the underlying eigenfunctions have a continuous and square-integrable second derivative. Recall that $\mathcal{D}\gamma = \gamma'(t)$ is the differential operator and define the roughness penalty by $\Psi(\gamma) = ||\mathcal{D}^2\gamma||^2$. Moreover, suppose that γ_m has square-integrable derivatives up to the degree four and that the second and the third derivative satisfy one of the following conditions:

1. $\mathcal{D}^2\gamma$, $\mathcal{D}^3\gamma$ are zero at the ends of the interval J
2. the periodicity boundary conditions of γ,$\mathcal{D}\gamma$, $\mathcal{D}^2\gamma$ and $\mathcal{D}^3\gamma$ on J.

The roughness penalty can be then written in the following way:

$$\begin{aligned} ||\mathcal{D}^2\gamma||^2 &= \int \mathcal{D}^2\gamma(s)\mathcal{D}^2\gamma(s)ds \\ &= \mathcal{D}\gamma(u)\mathcal{D}^2\gamma(u) - \mathcal{D}\gamma(d)\mathcal{D}^2\gamma(d) - \int \mathcal{D}\gamma(s)\mathcal{D}^3\gamma(s)ds \quad (4.18) \\ &= \gamma(u)\mathcal{D}^3\gamma(u) - \gamma(d)\mathcal{D}^3\gamma(d) - \int \gamma(s)\mathcal{D}^4\gamma(s)ds \quad (4.19) \\ &= \langle\gamma, \mathcal{D}^4\gamma\rangle, \quad (4.20) \end{aligned}$$

where d and u are the boundaries of the interval J and the first two elements in (4.18) and (4.19) are both zero under both conditions mentioned above.

Given a principal component function γ, with norm $||\gamma||^2 = 1$, the sample variance of the principal component can be penalized by dividing it by $1 + \alpha\langle\gamma, \mathcal{D}^4\gamma\rangle$:

$$PCAPV = \frac{\int\int \gamma(s)\widehat{\sigma}(s,t)\gamma(t)dsdt}{\int \gamma(t)(\mathcal{I} + \alpha\mathcal{D}^4)\gamma(t)dt}, \quad (4.21)$$

where $\mathcal{I}$ denotes the identity operator. The maximum of the penalized sample variance of the principal component (PCAPV) is an eigenfunction γ corresponding to the largest eigenvalue of the generalized eigenequation:

$$\int \widehat{\sigma}(s,t)\gamma(t)dt = \lambda(\mathcal{I} + \alpha\mathcal{D}^4)\gamma(s). \tag{4.22}$$

The smoothed principal components, solutions of (4.22) will be denoted in the following sections by $\hat{\gamma}_{\alpha,r}$ sorted w.r.t. the corresponding eingenvalues $\hat{\lambda}_{\alpha,r}$ of (4.22).

As already mentioned above, the resulting weight functions are no longer orthonormal in the L^2 sense. Since the weight functions are used as smoothed estimators of principal components functions, we need to rescale them so that they satisfy $||\hat{\gamma}_{\alpha,r}||^2 = 1$. The weight functions $\hat{\gamma}_{\alpha,r}$ can be interpreted as orthogonal in modified scalar product of the (generalized) Sobolev type

$$(f,g) = \langle f,g\rangle + \alpha\langle\mathcal{D}^2 f, \mathcal{D}^2 g\rangle. \tag{4.23}$$

The consistency results are given in following theorem:

THEOREM 4.2 *Assume* $\lambda_1 > \lambda_2 > ... > 0$ *and that the* $||\mathcal{D}^2\gamma_r|| < \infty$, $\alpha \to 0$ *as* $n \to \infty$ *and usual regularity conditions on* Γ *hold. Then for each* r

$$\hat{\lambda}_{\alpha,r} \to \hat{\lambda}_r \text{ for } n \to \infty$$

and

$$||\frac{\hat{\gamma}_{\alpha,r}}{||\hat{\gamma}_{\alpha,r}||} - \gamma_r|| \to 0 \tag{4.24}$$

Proof:
Complete proof based on the principle of complete induction and partially on the results of Dauxois et al. [1982] can be found in Silverman [1996]. Note that (4.24) holds by proper choice of sign of the eigenfunction. □

The method of Silverman [1996] is based on the choice of the norm (4.23). The method of Rice and Silverman [1991] is based on the more standard approach by penalizing the variance in the following way:

$$\int\int \gamma(s)\hat{\Gamma}_n\gamma(t)dsdt - \alpha_r||\mathcal{D}^2\gamma||^2$$

w.r.p. $\langle\gamma,\gamma_j\rangle = 0$ for $j < r$ and $\langle\gamma_r,\gamma_r\rangle = 1$, i.e. by penalizing the optimization criterion of FPCA with the roughness penalty in the way, usual in the e.g. smoothing (regression) splines. Rice and Silverman [1991] proposed to

decrease the amount of the penalization λ_r with increasing r. Pezzulli and Silverman [1993] gives the detailed theoretical study of this method, however Silverman [1996] suggest that the method of Rice and Silverman [1991] can be more advantageous and used under milder conditions.

Next alternative is to proceed in two steps – first smooth the functions X_i and then calculated the standard FPCA on the smoothed functions. Clearly the methods of Silverman [1996], Rice and Silverman [1991] combine the smoothing with the FPCA idea and hence are smoothing "at the right place" since the aim is to smooth (and regulize) the principal components.

It is probably clear to all readers, that this paragraph describes only the theoretical part of the SFPCA, implicitly assuming that the X_i are directly observable. As argued above this is almost never the case, hence some estimation or interpolation method needs to be considered. For this purposes, the functional basis expansion technique seems to be a very convenient way.

Implementation Using Basis Expansion

The basic notation remains same as in the section 4.2.3.

Moreover, define $\mathbf{K}$ to be a matrix whose elements are $\langle D^2\theta_j, D^2\theta_k\rangle$. Then the generalized eigenequation (4.22) can be approximated by:

$$\mathbf{W}Cov(\mathbf{C})\mathbf{W}\mathbf{u} = \lambda(\mathbf{W} + \alpha\mathbf{K})\mathbf{u}. \tag{4.25}$$

By finding matrix $\mathbf{L}$ for that holds: $\mathbf{L}\mathbf{L}^\top = \mathbf{W} + \alpha\mathbf{K}$ and defining $\mathbf{S} = \mathbf{L}^{-1}$ (4.25) can be rewritten into:

$$\{\mathbf{S}\mathbf{W}Cov(\mathbf{C})\mathbf{W}\mathbf{S}^\top\}(\mathbf{L}^\top\mathbf{u}) = \lambda\mathbf{L}^\top\mathbf{u}.$$

Algorithm SFPCA

1. calculate $\mathbf{C}$ and $\mathbf{W}$
2. using Cholesky decomposition calculate $\mathbf{L}$ and their inverse $\mathbf{L}^{-1}$
3. use symmetrical matrix eigenvalue-eigenvector routine and obtain eigenvalues and eigenvectors ($\mathbf{u}$) of $\mathbf{S}\mathbf{W}Cov(\mathbf{C})\mathbf{W}\mathbf{S}^\top$
4. calculate $\mathbf{b} = \mathbf{L}^{-1}\mathbf{u}$
5. renormalize $\mathbf{b}$ with respect to matrix $\mathbf{W}$, so that $\mathbf{b}^\top\mathbf{W}\mathbf{b} = 1$

If we are looking at the first K eigenfunctions as the best empirical basis for the functional observations, we may also re-orthonormalize coefficients $\mathbf{b}_r$ with respect to matrix $\mathbf{W}$, using Gramm-Schmidt procedure. The $\hat{\gamma}_{\theta,\alpha,r} = \mathbf{b}_r^\top\theta$ is then estimate of $\gamma_{\alpha,r}$ (again $\hat{\gamma}_{\theta,\alpha,r}$ belongs to the r-th largest eigenvalue $\hat{\lambda}_{\theta,\alpha,r}$).

Choice of the Smoothing Parameter

The natural question that appears in SFPCA is the selection of the smoothing parameter α. A popular and intuitive solution to this problem is the Cross-Validation. Ramsay and Silverman [2005] proposed the Cross-Validation, based on the fact that the eigenfunctions $\gamma_1, \gamma_2, \ldots, \gamma_m$ can be seen as an optimal empirical basis. In order to keep notation simple, assume that the sample mean $1/n \sum_i^n X_i = 0$, this can be done simply by subtracting the sample mean from functions X_i. (Using functional basis we subtract the sample mean of the coefficients $\bar{\mathbf{c}}$ from the coefficients $\mathbf{c}_i$.) Let $\mathbf{G}$ be the scalar product matrix of the first m eigenfunctions, i.e. $\mathbf{G}_{ij} = \langle \gamma_i, \gamma_j \rangle$. For a function X, define the "part" of X orthogonal to the subspace spanned by the basis $\gamma_1, \gamma_2, \ldots, \gamma_m$ as

$$\zeta_m = X - \sum_{i=1}^{n} \sum_{j=1}^{m} (\mathbf{G}^{-1})_{ij} \langle \gamma_i, X \rangle \gamma_j.$$

Recall that for the SFPCA the eigenfunctions are no longer orthonormal, thus $\mathbf{G} \neq \mathbf{I}$. As a performance measure of the empirical basis $\gamma_1, \gamma_2, \ldots, \gamma_m$, the quantity $E||\zeta_m||^2$ can be used, and the value of the cross-validation criterion evaluated:

$$\mathrm{CV}(\alpha) = \sum_{m=1}^{\infty} \mathrm{CV}_m(\alpha) = \sum_{m=1}^{\infty} \sum_{i=1}^{n} ||\zeta_m^{[i]}(\alpha)||^2 \tag{4.26}$$

where $\zeta_m^{[i]}$ is the part of X_i orthogonal to the subspace spanned by estimates $\gamma_{\alpha,\theta,1}$, $\gamma_{\alpha,\theta,2}$, $\ldots$, $\gamma_{\alpha,\theta,m}$ estimated from the data set with excluded i-th observation. In the next step we will choose the α^{opt} such that:

$$\alpha^{opt} = \arg\min \mathrm{CV}(\alpha)$$

In practice, we will use just the truncated version of the first sum (4.26) – given the data set containing n functions, we are able to estimate only first $n - 1$ eigenfunctions, we may also expect that the eigenfunctions of high "order" are estimated with a large error.

General Algorithm

This algorithm is partially taken from Ramsay and Silverman [2005].

1. Center the data $X_{\theta,i}$ (subtract the sample mean).
2. For a given α calculate $\zeta_m^{[i]}$ without using the $i - th$ observation. Calculate $\mathrm{CV}(\alpha)$ by truncating it at m.
3. Minimize $\mathrm{CV}(\alpha)$ with respect to α.

Modified Algorithm and Algorithm Using Basis Expansion

Following algorithm is a slight modification of the general algorithm proposed by Ramsay and Silverman [2005]. In addition the case of functional basis approach is considered:

1. Center the data $X_{\theta,i}$, (subtract the sample mean),
   Set $\mathbf{C} = \mathbf{C} - n^{-1}\mathbf{C}^\top \mathbf{1}_n$.

2. For a given α calculate the eigenfunctions $\hat{\gamma}_{\alpha,r}$ and eigenvalues $\hat{\lambda}_{\alpha,r}$, in general by solving (4.22), or $\hat{\gamma}_{\theta,\alpha,r}$, $\hat{\lambda}_{\theta,\alpha,r}$, $r = 1, \ldots, M_{max}$ employing the "Algorithm SFPCA" above, if using functional basis expansion.

3. Orthonormalize $\hat{\gamma}_{\theta,\alpha,r}$ $r = 1, \ldots, M_{max}$, e.g. by Gramm-Schmidt procedure:

   a) Set $\hat{\gamma}_{\theta,\alpha,r} := \hat{\gamma}_{\theta,\alpha,r} / \parallel \hat{\gamma}_{\theta,\alpha,r} \parallel$ i.e. $\mathbf{b}_r = \frac{\mathbf{b}_r}{(\mathbf{b}_r^\top \mathbf{W} \mathbf{b}_r)}$.

   b) for $j = r+1, \ldots, M_{max}$ set

   $$\hat{\gamma}_{\theta,\alpha,j} := \hat{\gamma}_{\theta,\alpha,j} - \langle \hat{\gamma}_{\theta,\alpha,r}, \hat{\gamma}_{\theta,\alpha,j} \rangle \hat{\gamma}_{\theta,\alpha,j}, i.e. \mathbf{b}_j = \mathbf{b}_j - (\mathbf{b}_r^\top \mathbf{W} \mathbf{b}_j)\mathbf{b}_j$$

   c) repeat a)–b) for $r = 1, \ldots M_{max}$.

4. Calculate the

   $$\zeta_{\theta,\alpha,m}^{[i]} = X_{\theta,i} - \sum_{j=1}^{m} \langle \gamma_{\alpha,\theta,j}, X_{\theta,i} \rangle \gamma_{\theta,\alpha,j} = \mathbf{c}_i^\top \boldsymbol{\theta} - \sum_{j=1}^{m} (\mathbf{b}_j^\top \mathbf{W} \mathbf{c}_i) \mathbf{c}_i^\top \boldsymbol{\theta}$$

   for $i = 1, \ldots, n$.

5. Calculate a truncated version of CV(α):

   $$\mathrm{CV}(\alpha) = \sum_{m=1}^{M_{max}} \mathrm{CV}_m(\alpha) = \sum_{m=1}^{M_{max}} \sum_{i=1}^{n} ||\zeta_{(\theta,\alpha,m)}^{[i]}||^2$$

   by truncating at m.

6. Minimize CV(α) with respect to α.

The advantage of this algorithm is that it is fully data driven. However, it is known from practice that the Cross Validation leads to unstable results. Another disadvantage of this method are high computational costs.

In this chapter we have presented the case with roughness penalty $\| \mathcal{D}^2\gamma \|$ similarly we could consider a more general case with roughness penalty $\| \mathcal{L}\gamma \|$.

Next practical question is how to optimize the CV(α). A natural answer is to use some iterative numerical procedure. However, as argued for example by Ramsay and Silverman [2005] a simple grid search will usually be enough to obtain practically suitable estimate of optimal α, since small changes in α do not make a big difference in resulting eigenfunctions.

Last topic that is of big importance in practice is the choice of L – number of functions used in the truncated KL-decomposition (4.11).

In practice many rule-of-thumb selecting procedures are used, e.g. selecting L so that the ratio of variance explained by the first L functions, equal to sum of first L largest eingenvalues to total variance (sum of all eigenvalue) is beyond the predefined threshold value e.g. 95%. Or the well known "elbow"-rule based on the plot of the eigenvalues.

Clearly if $\lambda_r = 0$ for $r \geq R$ then $R \leq L$ is naturally required. For gaussian case test based on asymptotic distribution of eigenvalues (for classical FPCA) can be based on the theorem 4.1.

An general automated (data-driven) procedure is not yet completely solved, a bootstrap method for the case where X_i are probability density functions was proposed by Kneip and Utikal [2001], see also the discussion in section 4.4.

4.2.5 Yield Curves Analysis

In this section we present an application of the methods described above to the analysis of the yield curve dynamics (evolution of the yield curves over time). First the EURIBOR-yield curve is considered. As announced in the section 3.5, the EURIBOR yield curve is a bench-mark interest rate published on the daily basis for a set of maturities 1 weeks, 2 weeks, 3 weeks, 1 month, 2 month, ... 12 months. Understanding of the time developement of the EURIBOR yield curve is essential for trading the interest rate derivatives that depends on the developement of EURIBOR such as already mentioned interest rate swaps.

First the discrete observations of the EURIBOR have to be represented by (smooth) functions. The EURIBOR interest rate can be assumed to be observed without (with negligibly small) error (it is already average of individual interest rates referred by the panel banks, see section 3.5.1 for more details).

A B-Spline basis with cubic B-Splines has been used, the knots sequence is equal to the observation grid, i.e. 1 weeks, 2 weeks, 3 weeks, 1 month, 2 month, ... 12 months. Clearly, this basis is able to interpolate the observed (discrete) data. However this interpolation can be rough between the knots see right part of the figure 4.1. Of course there is no reason to believe that yield curve behaves in this way, and as mentioned in section 3.5.1, the yield curve is assumed to be smooth.

In order to penalize the highly fluctuating interpolation the final coefficients have been obtained by minimizing the penalized least squares criterion (3.18), i.e.

$$\sum_{j=1}^{T} \left(R_{i,\tau_j} - \mathbf{c_i}^\top \boldsymbol{\theta}(\tau_j)\right)^2 + \alpha \parallel \mathcal{L}(\mathbf{c_i}^\top \boldsymbol{\theta}) \parallel^2 \tag{4.27}$$

where R_{i,τ_j} are the observed EURIBOR rates on the day i at the maturity τ_j, T is the number of the observed EURIBOR interest rates, since these interest rates are published for 1 weeks, 2 weeks, 3 weeks, 1 month, 2 month, ... 12 months, i.e. $T = 15$. The EURIBOR yield curve on the day i, $r_i(\tau)$, is approximated by $\mathbf{c_i}^\top \boldsymbol{\theta}(\tau)$. $\mathcal{L}$ is a linear differential operator,

$$\mathcal{L}(f) = (1 + \tau)\mathcal{D}^2(f)(\tau),$$

for any function f with $\parallel \mathcal{L}f \parallel < \infty$. The idea behind this choice is motivated by the figure 3.1 – the yield curve is flat for higher and more skewed for shorter maturities, this as a common phenomenon for all types of yield curves (except "flat"-constant yield curve). Obviously the $\mathcal{L}$ is penalizing the short maturity of the yield curve less than the long maturities. The smoothing parameter α

is set to be very small 1.10^{-9}. The aim is not to smooth the data but more or less "regularize" the interpolation.

Comparison of the penalized and standard approximation of the coefficients is plotted on the figure 4.9.

The construction of the smoothing operator $\mathcal{L}$ can be optimized in the similar way as the choice of the penalizing parameter α. A more sophisticated approach, where this operator is chosen from a broader class by a data driven procedure is commented in Ramsay and Silverman [2005].

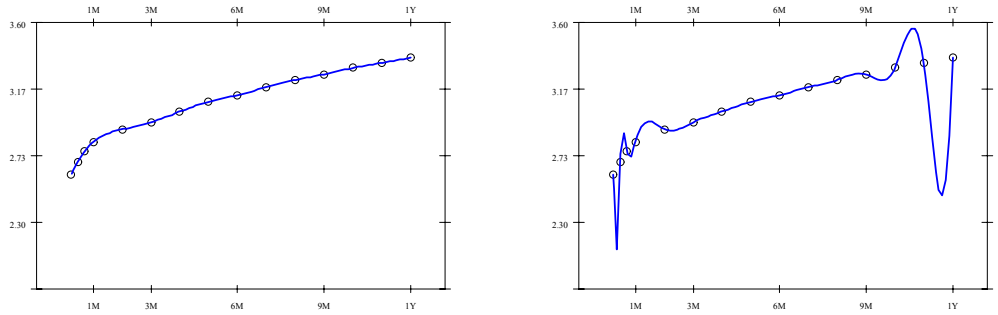

Figure 4.1: Example of the approximation of the EURIBOR yield curve using penalized cubic B-Splines, on June, 06 2006. In the left figure a regularized approximation, in the right figure, standard LS based approximation is plotted.

Figure 4.2 presents the penalized interpolation of the yield curves for the time period from January 1st 2004 to June 15, 2006. It can be observed that the yield curves are strongly moving across time. In the resent literature a variety of models for yield curve dynamics have been proposed. Focusing on the factor type of models, first and popular model is the parametric model introduced by Nelson and Siegel [1987] recently re-parametrized and simplified by Diebold and Li [2005], it is a three factor model of level, curvature and slope. Next broadly accepted model is model based on the PCA analysis, see Knez et al. [1994] originally proposed for analyzing the bond returns by multivariate PCA, among others, Blaskowitz et al. [2005] propose recently a method for forcasting the EURIBOR swap rate curves by combination of PCA and auto regressive models.

The main aim of this study is to investigate a factor model based on the functional principal components. The step from the multivariate to the functional principal components has a clear advantage, that using the function basis expansion the model can be evaluated for arbitrary point (i.e. for any

time to maturity). This is of great importance e.g. as already mentioned by discussing the calculation of IV, where the whole yield curve is needed. Next advantage of the functional PCA is that it can incorporate the functionals of the investigated datas, e.g. the derivative of the yield curve. The derivative of the yield curve has not much meaning in the multivariate PCA however can be directly addressed in FPCA. Derivative of the yield curve is connected to the forward curve (in fact the shape of the forward curve is completely determined by the derivative of the yield curve) and has a direct economic meaning, see (3.30).

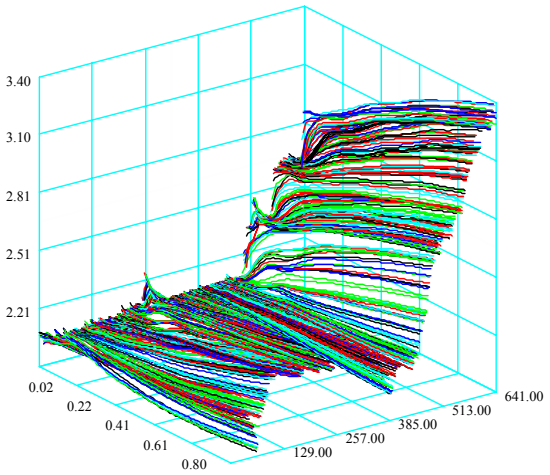

Figure 4.2: Time development of the EURIBOR yield curve. Each function (color line) corresponds to a yield curve on a certain day.

Motivating by the former multivariate PCA methods a L-factor linear return generating process for returns is assumed, $\Delta r_i(\tau) = r_i(\tau) - r_{i-1}(\tau)$, where $r_i(\tau)$ stands for the EURIBOR yield curve on a day i:

$$\Delta r_i = \mu + \sum_{r=1}^{L} \beta_{ri} v_r. \tag{4.28}$$

Assuming the orthonormality of v_r (4.28) is an analogue of the truncated KL decomposition (4.11) and v_r can be estimated in terms of FPCA - i.e. by eigenfunctions of the empirical covariance operator.

The returns Δr_i are plotted on the figure 4.3. Each function (color line) corresponds to a return on a certain day, the returns are ordered w.r.t. days

in the left (3D figure), the thick red line represents the sample mean function and two tick black lines are the variance bounds constructed as a mean plus or minus two times standard deviation. These variance bounds have only a visualization role (assuming an i.i.d. gaussian random sample the bounds represent the well known approximation of the 95% confidence intervals). It can be seen that the returns are centered around zero line and have higher variance for very short maturities up to 1 Month maturity afterwards the variance is increasing.

The figure 4.3 might be considered as "overprinted" – too many functions are plotted at one figure (especially around zero axis), in order to reduce the overprinting of the figure, the figure 4.4 presents an analogue of well known boxplot used in the univariate or multivariate analysis.

For each point (here the 100 equidistant grid points on $[0, 1]$ are used) a box-plot is constructed – the red point corresponds to the mean, three black functions correspond to the 25%, 50% and 75% quantiles (lower quartile, median and upper quartile), two blue lines are analogues of boxplot's whiskers constructed as 25% quantile $-2.5.IQR$ (lower line), IQR stands for Inter-Quartile-Range (difference between 25% and 75% quantile) and 75% quantile $+2.5.IQR$ (upper line). The dashed lines represent the functions exceeding the whiskers. Note that in the univariate situation the whiskers are typically plotted at the quartile $+/-1.5.IQR$ however in the (local) boxplots proposed here these should be set up according to the data, otherwise again an overprinting for whiskers exceedings may occur.

It can be seen that the median and mean functions are overlapping and except for the small range around 0.2 are in the center of two quartiles (and whiskers) lines, hence it may be concluded that the distribution is close to symmetric, centered around zero for all arguments (maturities).

Figures 4.5 and 4.6 visualize the covariance and correlation surfaces respectively. The left plots are the contour plots, each line corresponds to the "contour lines" with constant functional values at the level $i/17.range$ $i = 1, \ldots, 16$ of the covariance surface, see Härdle et al. [2000] for details.

Investigating the covariance and correlation surfaces, it is visible that the correlation in yield curves for higher maturities is very high and hence the parallel shift is probably the most important factor of the variation of the returns.

Finally the eigenfunctions $\hat{\gamma}_{\theta,r}, r = 1, \ldots, 4$ calculated from the empirical covariance operator (approximated by the basic expansion) are plotted in the left part of the figure 4.7. The corresponding eigenvalues (ordered) are plotted in the right part of the figure, in addition these values are summarized in the table 4.1.

The eigenanalysis has been performed using the smoothed (regularized)

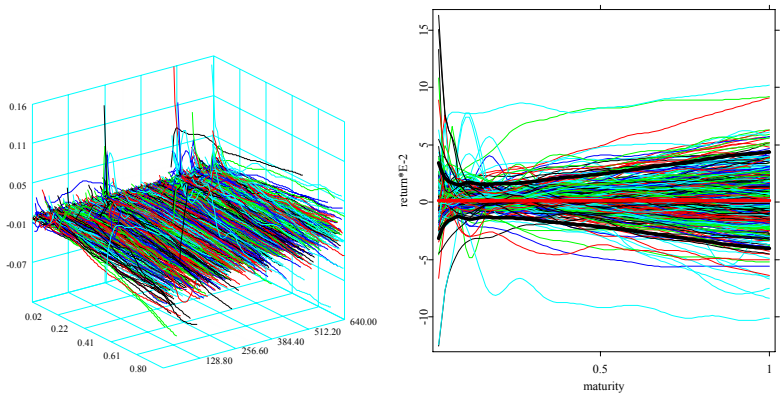

Figure 4.3: Returns on EURIBOR yield curves, each function (color line) corresponds to a return on a certain day, the returns are ordered w.r.t. to days in the left (3D figure), the thick red line represents the sample mean function and two tick black lines represents the "2σ bands".

FPCA approach described in the section 4.2.4. The smoothing parameter α in (4.21) was set by the CV criterion as presented in section 4.2.4, the CV criterion was minimized on the grid $1.10^{-9}.1.2^{l}, l = 1, \ldots, 30$ and the minimum was obtained at $1.71.10^{-8}$.

First eigenfunction (black line), that explains 88.6% of the sample variance, is positive for all values and hence can be roughly interpreted as the level shift, however the function is weighting longer maturities higher than the shorter, this interpretation coincide with the results concluded from the figures 4.6.

The second function (blue line), (6.1% explained variance) can be interpreted as the slope factor of the yield curve.

The third function (green line), (2.7% explained variance) is interpreted as curvature. In addition the short term effect is visible, this finding is in coincidence with the high variance for the short maturities.

The first three functions explain more than 95% of the sample variance. The number of functions L considered in the factor model (4.28) is set to $L = 3$.

Doing the same exercise for maturities from 2 to 30 years (swap based zero yield curves for period Januar 2004 – 15th Jun 2006, see figure 3.2 and enclosed discussion, first three eigenfunctions (ordered w.r.t. to corresponding eigenvales) as plotted in the figure 4.8 are obtained. The corresponding eigenvalues are plotted in the right part of the figure. These functions can

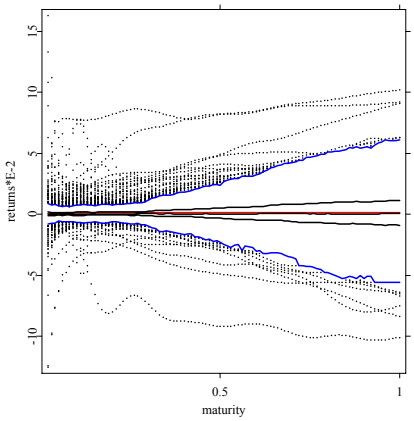

Figure 4.4: (Local) boxplots for returns EURIBOR yield curves.

	var. explained	cum. var. explained
$\hat{\gamma}_1$	0.886	0.886
$\hat{\gamma}_2$	0.061	0.947
$\hat{\gamma}_3$	0.022	0.970
$\hat{\gamma}_4$	0.011	0.979

Table 4.1: Variance explained by the eigenfunctions of sample of EURIBOR returns. First column are the variances explained by each eigenfunction ($\nu_r = \hat{\lambda}_r / \sum_j \hat{\lambda}_j$) in the second column the cumulative sum of explained variances are listed ($\sum_{j=1}^{r} \nu_j$).

be viewed using standard interpretation as level (black line, 87.1% explained variance), slope (blue line, 7.1% explained variance) and curvature effect (green line, 1.4% explained variance).

The Smoothed FPCA technique is used, the CV is minimized over same grid as for EURIBOR yield curve, the minimum is achieved at $1.8488.10^{-8}$. The third eigenfunction can be seen as undersmoothed, this is partially caused by rough (swap) yield curves in 2004 in this region.

Concluding, this small investigation study shows the application of the functional data analysis by the analysis of the dynamics of yield curves. First advantage of this approach, in compare to standard multivariate methods, is that the model can be evaluated on arbitrary fine grid.

Moreover, the study shows that the short maturities seem to have its "own life", a factor capturing the maturities shorter than 1 month was identified, the factors for larger maturities have the well known level, slope and

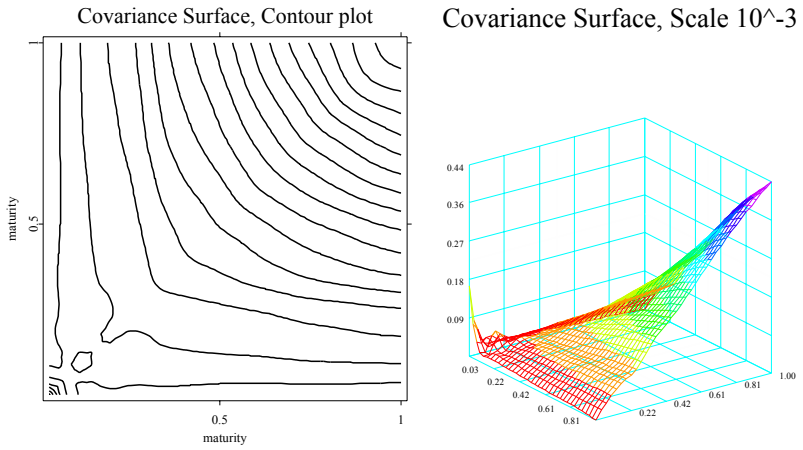

Figure 4.5: Covariance surface plot (right figure) and contour plot of the covariance function for EURIBOR returns (left figure).

	var. explained	cum. var. explained
$\hat{\gamma}_1$	0.871	0.871
$\hat{\gamma}_2$	0.071	0.941
$\hat{\gamma}_3$	0.014	0.956

Table 4.2: Variance explained by the eigenfunctions of sample of returns on yield curves based on swap rates. First column are the variances explained by each eigenfunction ($\nu_r = \hat{\lambda}_r / \sum_j \hat{\lambda}_j$). In the second column the cumulative sum of explained variances are listed ($\sum_{j=1}^{r} \nu_j$).

curvature interpretation.

Further comments Two comments that are beyond the scope of this study, but are worth to be mentioned here: first in this study we use essentially the smoothness of the yield curve, the linear differential operator is giving the opportunity to penalize the smoothness in different regions differently (allowing for local behavior) by low computational costs.

Another choices of linear differential operator can be very promising, e.g. a simple linear differential operator performed on the zero yield curve results in the (instantaneous) forward yield curve, see (3.30), and functional "qualities" of forward yield curves like smoothness, positivity or any other based on some economic theory can be translated into proper operator and further into proper estimation technique.

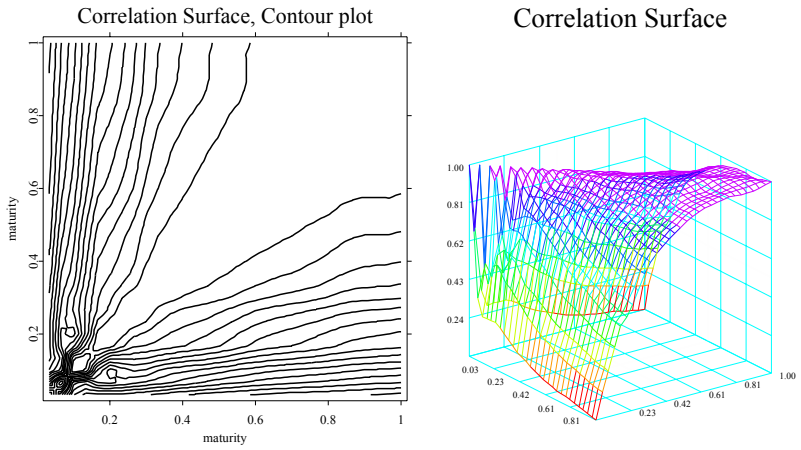

Figure 4.6: Correlation surface plot (right figure) and contour plot of the covariance function for EURIBOR returns (left figure).

Secondly the next step of application of the factor model (4.28) is the analysis of the factor loading β_{ri} (or its estimates). This is typically done by econometric time-series analysis, most frequently a vector auto regressive models are used, see Diebold and Li [2005] among others. The time-series analysis proceed as if the factor loadings β_{ri} would be estimated without error, assuming this, the time-serie model can be chosen by best prediction argument from a broad class of models. Blaskowitz et al. [2005] among others propose a combination of the econometric arguments and parameter choice of the estimation method based on best prediction arguments. This argument can be also very useful here. Clearly the estimators $\hat{\beta}_{ri} = \langle \hat{\gamma}_{\theta,\alpha,r}, X_{\theta_i} \rangle$ are depending on the tuning parameters α, θ and L hence the minimizing of the prediction error can be done not only w.r.t. to time series model but also w.r.t. to all parameters involved in estimation. Of course this approach is connected with high computational costs, however it gives a natural and promising alternative.

4.2.6 Dual Approach

The following two sections follow closely the argumentation presented in Benko et al. [2006b], where the method was originally proposed. Their approach is motivated by the duality relation between row and column spaces of a data matrix – (4.4) and (4.5). The elements of this matrix, $X_l^\top X_k = \langle X_l, X_k \rangle$, in the multivariate case, for $i, j = 1, \ldots, n$. As already argued, in the functional case the notation remains same, only the scalar product stands

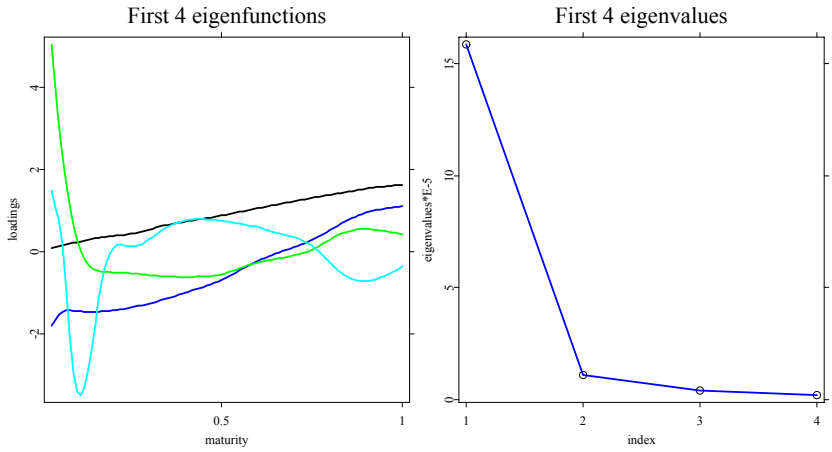

Figure 4.7: First four eigenfunctions and corresponding eigenvalues of sample of EURIBOR returns.

for the L^2 scalar product, hence the object of interest is the matrix of scalar products:

$$M_{lk} = \langle X_l - \bar{X}, X_k - \bar{X} \rangle, \quad l, k = 1, \ldots, n. \tag{4.29}$$

In the same way as in multivariate case, some simple linear algebra shows that all nonzero eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \ldots$ of $\hat{\Gamma}_n$ and $l_1 \geq l_2 \ldots$ of M are related by $\hat{\lambda}_r = l_r/n$, $r = 1, 2, \ldots$. When using the corresponding orthonormal eigenvectors $p_1, p_2, \ldots$ of M, the empirical scores $\hat{\beta}_{ri}$ as well as the empirical eigenfunctions $\hat{\gamma}_r$ are obtained by $\hat{\beta}_{ri} = \sqrt{l_r} p_{ir}$ and

$$\hat{\gamma}_r = \frac{1}{\sqrt{l_r}} \sum_{i=1}^{n} p_{ir} \left(X_i - \bar{X} \right) = \frac{1}{\sqrt{l_r}} \sum_{i=1}^{n} p_{ir} X_i. \tag{4.30}$$

In practice the functions are typically not directly observed and the following model is assumed:

$$Y_{ik} = X_i(t_{ik}) + \varepsilon_{ik}, \quad k = 1, \ldots, T_i, \tag{4.31}$$

where ε_{ik} are independent noise terms with $E(\varepsilon_{ik}) = 0$, $Var(\varepsilon_{ik}) = \sigma_i^2$.

In the simplest case, where the functions are observed on the balanced, equidistant design $t_{ij} = j/T$, $i, j = 1, \ldots n$, $T = T_i, i = 1, \ldots, n$, then M can be estimated by:

$$\widehat{M_{ij}} = T^{-1} \sum_{k=1}^{T} (Y_{ik} - \bar{Y}_{\cdot k})(Y_{jk} - \bar{Y}_{\cdot k}), \; i \neq j,$$

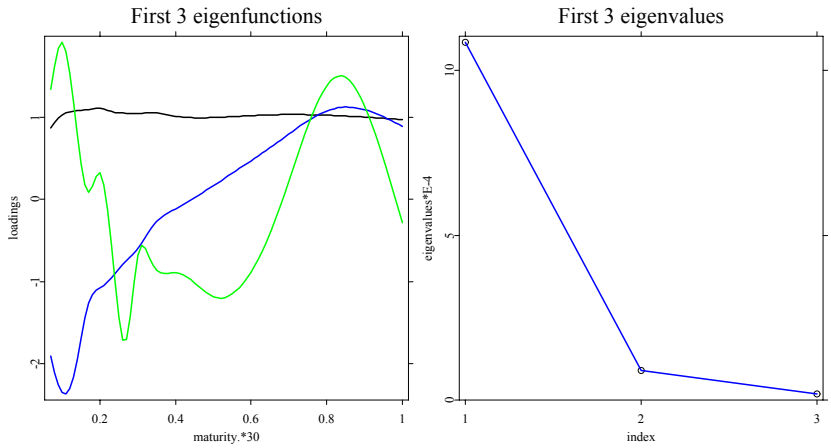

Figure 4.8: First three eigenfunctions and corresponding eigenvalues of sample of returns on yield curves based on Swap rates the X axis is the maturity and scaled by 30, i.e. 1 denotes the maturity of 30 years.

and

$$\widehat{M_{ii}} = T^{-1} \sum_{k=1}^{T} (Y_{ik} - \bar{Y}_{\cdot k})^2 - \hat{\sigma}_i^2, \text{ for } i \neq j,$$

where $\hat{\sigma}_i^2$ denotes estimator of variance σ_i and $\bar{Y}_{\cdot k} = n^{-1} \sum_i Y_{ik}$.

In the case of a random and design, Benko et al. [2006b] proposed a following adjustment: define the ordered sample

$$t_{i(1)} \leq t_{i(2)} \leq \cdots \leq t_{i(T_i)}$$

of design points, and for $j = 1, \ldots, T_i$. Let $Y_{i(j)}$ denote the observation belonging to $t_{i(j)}$, with $t_{i(0)} = -t_{i(1)}$ and $t_{i(T_i+1)} = 2 - t_{i(T_i)}$ set

$$\chi_i(t) = \sum_{j=1}^{T_i} Y_{i(j)} I\left(t \in \left[\frac{t_{i(j-1)} + t_{i(j)}}{2}, \frac{t_{i(j)} + t_{i(j+1)}}{2} \right] \right), \ t \in [0,1],$$

where $I(\cdot)$ denotes the indicator function. For $i \neq j$ define an estimate of M_{ij} by

$$\widehat{M_{ij}} = \int_0^1 \{\chi_i(t) - \bar{\chi}(t)\} \{\chi_j(t) - \bar{\chi}(t)\} \, dt,$$

where $\bar{\chi}(t) = n^{-1} \sum_{i=1}^{n} \chi_i(t)$. By redefining $t_{i(1)} = -t_{i(2)}$ and $t_{i(T_i+1)} = 2 - i(T_i)$, set

$$\chi_i^*(t) = \sum_{j=2}^{T_i} Y_{i(j-1)} I\left(t \in [\frac{t_{i(j-1)} + t_{i(j)}}{2}, \frac{t_{i(j)} + t_{i(j+1)}}{2}] \right),$$

$t \in [0,1]$. The estimators of the diagonal terms M_{ii} is constructed as

$$\widehat{M_{ii}} = \int_0^1 \{\chi_i(t) - \bar{\chi}(t)\} \{\chi_i^*(t) - \bar{\chi}(t)\} \, dt \tag{4.32}$$

The aim of using the estimator (4.32) for the diagonal terms is to avoid the additional bias implied by $E_\varepsilon(Y_{ik}^2) = X_i(t_{ij})^2 + \sigma_i^2$. Here E_ε denotes conditional expectation given t_{ij}, X_i. Alternatively a bias corrected estimator can be constructed using some nonparametric estimation of variance σ_i^2, e.g. the difference based model-free variance estimators studied in Hall et al. [1990] can be employed.

The estimate defined above is based on an approximation of the functions X_i by a piecewise constant function χ_i, a simple case of a linear smoother. Clearly the estimation method of the scalar-product matrix M, proposed above can be generalized using any type of linear smoother of X_i. Let $\mathbf{Y}_i \stackrel{\text{def}}{=} (Y_{i1}, \ldots, Y_{iT_i})$, $\mathbf{X}_i \stackrel{\text{def}}{=} \{X_i(t_{i1}), \ldots, X_i(t_{iT_i})\}$. Consider a local linear estimator:

$$\widetilde{X}_i(t) = \mathbf{w}_i(t)^\top \mathbf{Y}_i, \ t \in [0,1], \ i = 1, \ldots, n,$$

where $\mathbf{w}_i(t)$ is vector of (local) weights. A plug-in estimator of the M has the following form:

$$\hat{M}_{ij} = \mathbf{Y}_i^\top \mathbf{A}(i,j) \mathbf{Y}_j, \tag{4.33}$$

where $\mathbf{A}$ is the matrix of elements $\mathbf{A}(i,j)_{k,l} = \int_0^1 w_{i,k}(t) w_{j,k}(t) dt$. Since

$$E\left[\hat{M}_{ij} | \mathbf{t}_i, \mathbf{t}_j\right] = \mathbf{X}_i^\top \mathbf{A}(i,j) \mathbf{X}_j + I(i=j) \mathrm{tr}\left\{\mathbf{A}(i,j)\right\},$$

where $\mathbf{t}_i \stackrel{\text{def}}{=} (t_{i1}, \ldots, t_{iT_i})$, the estimators of the diagonal terms can be adjusted by "bias-corrected" estimator:

$$\hat{M}_{ij} = \mathbf{Y}_i^\top \mathbf{A}(i,j) \mathbf{Y}_j - I(i=j) \hat{\sigma}^2 \mathrm{tr}\left\{\mathbf{A}(i,j)\right\}. \tag{4.34}$$

$\hat{\sigma}^2$ is a nonparametric estimator of the residual variance σ_i^2, e.g. the difference based model-free variance estimators studied in Hall et al. [1990] can be employed again. A deeper study of the estimation of the quadratic regression functionals (diagonal terms M_{ii}, essentially based on the local polynomial estimation, can be found in Fan and Huang [1999]. In case of common design ($t_{il} = t_{jl}, i, j = 1, \ldots k, l = 1, \ldots T$) the theoretical results can be directly applied also to the non-diagonal terms M_{ij}. The adjustment of these estimators to the non-common designs is however difficult and computationaly very intensive, hence unappropriate for this particular application.

The eigenvalues $\hat{l}_1 \geq \hat{l}_2 \ldots$ and eigenvectors $p_1, p_2, \ldots$ of the resulting matrix $\widehat{M}$ then estimate $\hat{\lambda}_{r;T} = \hat{l}_r / n$ and $\hat{\beta}_{ri;T} = \sqrt{\hat{l}_r} \hat{p}_{ir}$ of $\hat{\lambda}_r$ and $\hat{\beta}_{ri}$. The

only unknown term in (4.30) are then X_i. X_i – the true unknown function can be estimated by nonparametric estimate $\hat{X}_i$ (e.g. local polynomial with bandwidth b). This yield the estimate $\hat{\gamma}_{r;T}$:

$$\hat{\gamma}_{r;T} = \frac{1}{\sqrt{\hat{l}_r}} \sum_{i=1}^{n} \hat{p}_{ir} \hat{X}_i. \tag{4.35}$$

The following theorem quantifies the magnitude of the error caused by the estimation.

THEOREM 4.3 *Assume that X_i is a.s. twice continuously differentiable. There exists a constant $D_1 < \infty$ such that the derivatives are bounded by $\sup_t E[X_i^{'}(t)^4] \leq D_1$ as well as $\sup_t E[X_i^{''}(t)^4] \leq D_1$. The design points t_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, T_i$ are i.i.d random variables which are independent of X_i and ε_{ij}. The design density f is continuous on $[0, 1]$ and satisfies $\inf_{t\in[0,1]} f(t) > 0$. For any i the error terms ε_{ik} are i.i.d. zero mean random variables with $Var(\varepsilon_{ik}) = \sigma_i^2$. Furthermore, ε_{ik} is independent of X_i, and there exists a constant D_2 such that $E(\varepsilon_{ik}^8) < D_2$ for all i, n. The estimates $\hat{X}_i$ used in (4.35) are determined by either a local linear or a Nadaraya-Watson kernel estimator with smoothing parameter b and kernel function K. K is a continuous probability density which is symmetric at 0. Furthermore assume that the conditions of theorem 4.1 are fulfilled and that $\inf_{s\neq r} |\lambda_r - \lambda_s| > 0$ holds for some $r = 1, 2, \ldots$. Then*

i) $n^{-1} \sum_{i=1}^{n} (\hat{\beta}_{ri} - \hat{\beta}_{ri;T})^2 = \mathcal{O}_p(T^{-1})$ *and*

$$|\hat{\lambda}_r - \frac{\hat{l}_r}{n}| = \mathcal{O}_p(T^{-1} + n^{-1}). \tag{4.36}$$

ii) If additionally $(Tb^2)^{-1} \to 0$ as $n, T \to \infty$, then for all $t \in [0, 1]$

$$|\hat{\gamma}_r(t) - \hat{\gamma}_{r;T}(t)| = \mathcal{O}_p\{b^2 + (nTb)^{-1/2} + (Tb^{1/2})^{-1} + n^{-1}\}. \tag{4.37}$$

Proof:
Proof is given in Benko et al. [2006b]. □

As already mentioned the theoretical results in functional data analysis are usually based on the implicit assumption that the additional error due to (4.31) is negligible, and that one can proceed "as if" the functions X_i were directly observed. Consider one of the following two cases:

1) T is much larger than n, i.e. $n/T^{4/5} \to 0$, and that the smoothing parameter b in (4.35) is of order $T^{-1/5}$ (optimal smoothing of individual functions), or

2) T is smaller than n but $n/T^2 \to 0$, and that an (undersmoothing) bandwidth $b \sim (nT)^{-1/5}$ is used.

In both cases Theorem 4.3 together with Theorem 4.1 imply that

$$|\hat{\lambda}_r - \frac{\hat{l}_r}{n}| = o_p(|\hat{\lambda}_r - \lambda_r|)$$

as well as

$$\|\hat{\gamma}_r - \hat{\gamma}_{r;T}\| = o_p(|\hat{\gamma}_r - \gamma_r|).$$

Hence the inference about functional principal components will be then first order equivalent to an inference based on known functions X_i.

The properties of the procedure are illustrated by small simulated example.

4.2.7 Example

For the illustration purposes, we use a simulated functional data set of random linear combinations of two Fourier functions:

$$X_i(t_{ik}) = \beta_{1i}\sqrt{2}\sin(2\pi t_{ik}) + \beta_{2i}\sqrt{2}\cos(2\pi t_{ik}) + \varepsilon_{ik}, \tag{4.38}$$

where the factor loadings are normally distributed with $\beta_{1i} \sim N(0, 6)$, $\beta_{2i} \sim N(0, 4)$, the error terms $\varepsilon_{ik} \sim N(0, 0.25)$ (all of them i.i.d. over i and k). The functions are generated ("observed") on the uniformly i.i.d. grid $t_{ik} \sim U[0, 1]$, $k = 1, \ldots, T = 150$, $i = 1, \ldots, n = 40$. The estimators $\hat{X}_i$ are obtained by the local constant (Nadaraya-Watson) estimator with Epanechnikov kernel and bandwidth $b = 0.07$.

Estimators $\hat{X}_i$ of the simulated functional data set and estimator of the first eigenfunction are displayed in the figure 4.9. The figure 4.11 gives another insight into the finite sample behavior. Here we have repeated the simulations 50 times, with $\beta_{1i} \sim N(0, 6)$, $\beta_{2i} \sim N(0, 4)$, $\varepsilon_{ik} \sim N(0, 0.25)$. We can see that the variation of the sample generated by the scheme (4.38) is essentially reflected in some shift of the estimated eigenfunction.

4.2.8 Choice of the Smoothing Parameter

The case 2) suggests the choice of the (undersmoothing) bandwidth b in the estimation of γ_r, this can be also motivated by (4.35) since $\hat{\gamma}_{r;T}$ is defined as a weighted average of all estimated sample functions. As argued by Benko et al. [2006b], averaging reduces variance, and efficient estimation of $\hat{\gamma}_r$ therefore requires *undersmoothing* of individual function estimates $\hat{X}_i$. With other

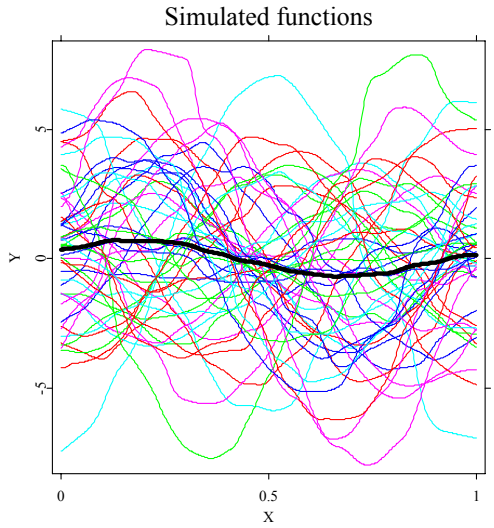

Figure 4.9: Simulated example, the Nadaraya-Watson estimators of simulated functions are plotted (b=0.07). Estimated mean functions (black thick line).

words, an optimal choice of a smoothing parameter $b \sim (nT)^{-1/5}$ yields the rate of convergence $\| \hat{\gamma}_r - \hat{\gamma}_{r;T} \| = \mathcal{O}_p\{(nT)^{-2/5}\}$. $|\hat{\lambda}_r - \frac{\hat{l}_r}{n}| = \mathcal{O}_p(T^{-1} + n^{-1})$. When using standard methods (based on direct eigenanalysis of empirical covariance operator of nonparametrically estimated curves $\hat{X}_i$) it does not seem to be possible to obtain a corresponding rate of convergence, since any smoothing bias $|E[\hat{X}_i(t)] - X_i(t)|$ will invariably affect the quality of the corresponding estimate of λ_r.

An automated choice of the smoothing parameters can be for instance based on a classical Cross-Validation argument, using the "best empirical basis"-feature of the eigenfunctions – for a fixed $s \in N$ let $\hat{\mu}_{T,-i}$ and $\hat{\gamma}_{r;T,-i}$, $r = 1, \ldots, s$ denote the estimates of $\hat{\mu}$ and $\hat{\gamma}_r$ obtained from the data (Y_{kj}, t_{kj}), $k = 1, \ldots, i-1, i+1, \ldots, n$, $j = 1, \ldots, T_k$. By (4.35) these estimates depend on b, and one may approximate an optimal smoothing parameter by minimizing

$$\sum_i \sum_j \left\{ Y_{ij} - \hat{\mu}_{T,-i}(t_{ij}) - \sum_{r=1}^{s} \hat{\vartheta}_{ri} \hat{\gamma}_{r;T,-i}(t_{ij}) \right\}^2$$

over b, where $\hat{\vartheta}_{ri}$ denote ordinary least squares estimates of $\hat{\beta}_{ri}$.

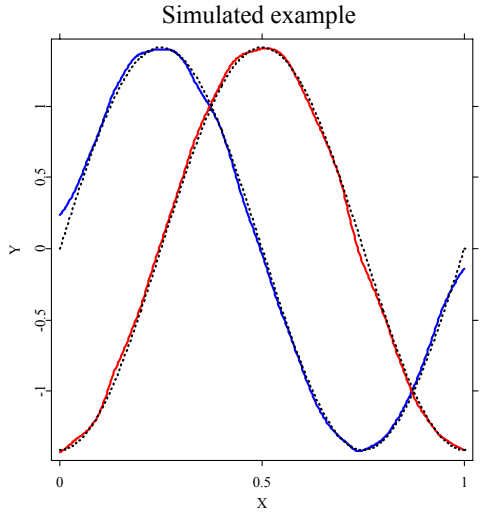

Figure 4.10: Simulated example, estimated first (blue) and second (red) eigenfunction, true eigenfunctions: (first blue, second red dashed).

There are two ways of improving this approach:

- Different bandwidths for different r. Since $< \gamma_s, \gamma_r >= 0$ for $s < r$, the number of zero crossings, peaks and valleys of γ_r has to increase with r. Hence, in tendency γ_r will be less and less smooth as r increases. At the same time, $\lambda_r \to 0$ which means that for large r the r-th eigenfunctions will only possess a very small influence on the structure of X_i. This in turn means that the relative importance of the error terms ε_{ij} in (4.31) on the structure of $\hat{\gamma}_{r;T}$ will increase with r. This arguments suggests that one should select different bandwidths b_r when estimating different (decreasing in r) functional principal components by (4.35).

- Different bandwidths for different i. In case of unbalanced design ($T_i \neq T_j$) a bandwidth b_i may be considered. The CV-criterion can be adjusted in this case by the following algorithm:

  1. Estimate the bandwidth h_i, e.g. optimal w.r.t. to ISE of $\hat{X}_i$, by a CV criterion $\sum (Y_{ij} - \hat{X}_{-j,i,h})^2$, where $\hat{X}_{-j,i,h}$ is "j"th-leave-one-out estimator of X_i and a global h (global over i) that minimize the sum of ISE for all i by sum over i of the CV criterions.
  2. Estimate the global bandwidth b by CV criterion described above.
  3. Set bandwidth $b_i = (h_i/h)b$.

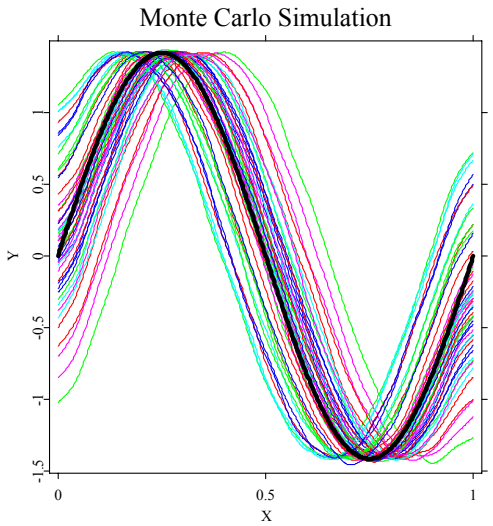

Figure 4.11: Monte Carlo Simulation, 50 replications, thin lines are estimated first eigenfunctions, the bold black line is the true eigenfunction

The idea is, that the individual properties of X_i that affect the choice of optimal bandwidth b_i, e.g. $||X''||$, T_i,... is already stored in h_i. However this approach is computationally highly involved.

Note, that the bias corrected estimator (4.32) may yield negative eigenvalues. In practice these values will be small and will have to be interpreted as zero. Furthermore, the eigenfunctions determined by (4.35) may not be exactly orthogonal. Again, when using reasonable bandwidths, this effect will be small, but certainly (4.35) may be followed by suitable orthogonalization procedure.

4.3 Two Sample Problem and FPCA

Main focus of this section is the two sample inference based on the FPCA. Denote two independent samples by

$$\begin{aligned} X_1^{(1)}, X_2^{(1)}, \ldots, X_{n_1}^{(1)} &\sim X^{(1)}, \\ X_1^{(2)}, X_2^{(2)}, \ldots, X_{n_2}^{(2)} &\sim X^{(2)}. \end{aligned}$$

The so-called "two-sample" problem is to test whether the distributions of the random variables $X^{(1)}$ and $X^{(2)}$ coincide. As argued in the previous

section the distributions of both random variables can be accessed by the corresponding KL-expansions:

$$X_i^{(p)} = \mu^{(p)} + \sum_{r=1}^{\infty} \beta_{ri}^{(p)} \gamma_r^{(p)}, \quad p = 1, 2, \tag{4.39}$$

where again $\gamma_r^{(p)}$ are the eigenfunctions of the respective covariance operator $\Gamma^{(p)}$ corresponding to the eigenvalues

$$\lambda_1^{(p)} = E\{(\beta_{1i}^{(p)})^2\} \geq \lambda_2^{(p)} = E\{(\beta_{2i}^{(p)})^2\} \geq \ldots, \ p = 1, 2.$$

The difference in the distributions are reflected in the differences of the KL-expansions. The interesting hypothesis, that are considered here are:

$$\begin{aligned} H_{0_1} &: \mu^{(1)} = \mu^{(2)} \\ H_{0_{2,r}} &: \gamma_r^{(1)} = \gamma_r^{(2)}, \ r = 1, 2, \ldots \end{aligned}$$

Hypothesis $H_{0_{2,r}}$ will be denoted by "common eigenfunctions" hypothesis. If $\gamma_r^{(1)} = \gamma_r^{(2)}$ is accepted then only the (scalar) factor loadings $\beta_{ri}^{(p)}$ may vary across samples. It can be seen as a functional generalization of the concept of "common principal components" as introduced by Flury [1988] in multivariate analysis.

The hypothesis of equality of eigenfunctions can be naturally relaxed to the hypothesis of the equality of the L-dimensional eigenspaces generated by the first L eigenfunctions. Therefore, let $\mathcal{E}_L^{(1)}$ and $\mathcal{E}_L^{(2)}$ denote the L-dimensional linear function spaces generated by the eigenfunctions $\gamma_1^{(1)}, \ldots,$ $\gamma_L^{(1)}$ and $\gamma_1^{(2)}, \ldots, \gamma_L^{(2)}$, respectively. We then aim to test the null hypothesis:

$$H_{0_{4,L}} : \mathcal{E}_L^{(1)} = \mathcal{E}_L^{(2)},$$

where $\mathcal{E}_L^{(p)} = \text{span}\{\gamma_1^{(p)}, \ldots, \gamma_L^{(p)}\}$.

Next, difference in the distributions of these two samples can appear if the eigenfunctions are equal i.e., $H_{0_{2,r}}$ is accepted, but the distributions $\beta_{ri}^{(p)}$, $p = 1, 2$ of the factor loadings differ. Recall that by definition $E\{\beta_{ri}^{(p)}\} = 0$, $E\{\beta_{ri}^{(p)}\}^2 = \lambda_r^{(p)}$, and $\beta_{si}^{(p)}$ is uncorrelated with $\beta_{ri}^{(p)}$ if $r \neq s$. In case of Gaussian (functional) random variables $X^{(1)}$, $X^{(2)}$, the $\beta_{ri}^{(p)}$ are also independent normally distributed $\beta_r^{(p)} \sim N(0, \lambda_r^{(p)})$ random variables. The test of equality of the distributions of $\beta_r^{(p)}$ then translates to

$$H_{0_{3,r}} : \lambda_r^{(1)} = \lambda_r^{(2)}, \ r = 1, 2, \ldots$$

One may think that a general distribution-free test may be based on the well-known Kolmogorov-Smirnoff test based on the $\beta_{ri}^{(1)}$ $i = 1, \ldots, n_1$ and $\beta_{ri}^{(2)}$ $i = 1, \ldots, n_1$. However the power of the Kolmogorov-Smirnoff test is known to be small and since even in the case of directly observable functions $X_i^{(p)}$ the estimation of the factor loadings is $\sqrt{n}_p$ consistent. Therefore, this test is not appropriate.

Test Statistics

Denote the sample mean of the p-th sample by

$$\hat{\mu}^{(p)}(t) = \frac{1}{n_p} \sum_{i=1}^{n_p} X_i^{(p)}(t),$$

and let

$$\hat{\lambda}_1^{(p)} \geq \hat{\lambda}_2^{(p)} \geq \ldots$$

and

$$\hat{\gamma}_1^{(p)}, \hat{\gamma}_2^{(p)} \geq \ldots$$

denote eigenvalues and corresponding eigenfunctions of the empirical covariance operator $\hat{\Gamma}_{n_p}^{(p)}$ of $X_1^{(p)}, X_2^{(p)}(t), \ldots, X_{n_p}^{(p)}$, for $p = 1, 2$. The following test statistics are defined in terms of $\hat{\mu}^{(p)}$, $\hat{\lambda}_r^{(p)}$ and $\hat{\gamma}_r^{(p)}$.

The tests of the hypotheses $H_{0_1}, H_{0_{2,r}}$ and $H_{0_{3,r}}$ rely on the squared distances of the corresponding functions:

$$\begin{aligned} D_1 &\stackrel{\text{def}}{=} \|\hat{\mu}^{(1)} - \hat{\mu}^{(2)}\|^2, \\ D_{2,r} &\stackrel{\text{def}}{=} \|\hat{\gamma}_r^{(1)} - \hat{\gamma}_r^{(2)}\|^2, \\ D_{3,r} &\stackrel{\text{def}}{=} |\hat{\lambda}_r^{(1)} - \hat{\lambda}_r^{(2)}|^2. \end{aligned}$$

Benko et al. [2006b] proposed the following test-procedure: the respective null-hypothesis has to be rejected if $D_1 \geq \Delta_{1;1-\alpha}$, $D_{2,r} \geq \Delta_{2,r;1-\alpha}$ or $D_{3,r} \geq \Delta_{3,r;1-\alpha}$, where $\Delta_{1;1-\alpha}$, $\Delta_{2,r;1-\alpha}$ and $\Delta_{3,r;1-\alpha}$ denote the critical values of the distributions of

$$\begin{aligned} \Delta_1 &\stackrel{\text{def}}{=} \|\hat{\mu}^{(1)} - \mu^{(1)} - (\hat{\mu}^{(2)} - \mu^{(2)})\|^2, \\ \Delta_{2,r} &\stackrel{\text{def}}{=} \|\hat{\gamma}_r^{(1)} - \gamma_r^{(1)} - (\hat{\gamma}_r^{(2)} - \gamma_r^{(2)})\|^2, \\ \Delta_{3,r} &\stackrel{\text{def}}{=} |\hat{\lambda}_r^{(1)} - \lambda_r^{(1)} - (\hat{\lambda}_r^{(2)} - \lambda_r^{(2)})|^2. \end{aligned}$$

Obviously, the distributions of the different Δ's cannot be accessed directly, since they depend on the unknown true population mean, eigenvalues and

eigenfunctions. However, it will be shown below that these distributions and hence their critical values are approximated by the bootstrap distribution of

$$\begin{aligned}\Delta_1^* &\stackrel{\text{def}}{=} \|\hat{\mu}^{(1)*} - \hat{\mu}^{(1)} - (\hat{\mu}^{(2)*} - \hat{\mu}^{(2)})\|^2,\\ \Delta_{2,r}^* &\stackrel{\text{def}}{=} \|\hat{\gamma}_r^{(1)*} - \hat{\gamma}_r^{(1)} - (\hat{\gamma}_r^{(2)*} - \hat{\gamma}_r^{(2)})\|^2,\\ \Delta_{3,r}^* &\stackrel{\text{def}}{=} |\hat{\lambda}_r^{(1)*} - \hat{\lambda}_r^{(1)} - (\hat{\lambda}_r^{(2)*} - \hat{\lambda}_r^{(2)})|^2.\end{aligned}$$

where $\hat{\mu}^{(1)*}$, $\hat{\gamma}_r^{(1)*}$, $\hat{\lambda}_r^{(1)*}$ as well as $\hat{\mu}^{(2)*}$, $\hat{\gamma}_r^{(2)*}$, $\hat{\lambda}_r^{(2)*}$ are estimates to be obtained from independent bootstrap samples $X_1^{1*}(t), X_2^{1*}(t), \ldots, X_{n_1}^{1*}(t)$ as well as $X_1^{2*}(t), X_2^{2*}(t), \ldots, X_{n_2}^{2*}(t)$.

This test procedure is motivated by the following insights:

1) Under each of our null-hypotheses the respective test statistics D is equal to the corresponding Δ. The test will thus asymptotically possess the correct level: $P(D > \Delta_{1-\alpha}) \approx \alpha$.

2) If the null hypothesis is false, then $D \neq \Delta$. Compared to the distribution of Δ the distribution of D is shifted by the difference in the true means, eigenfunctions, or eigenvalues. In tendency D will be larger than $\Delta_{1-\alpha}$.

Certainly, $H_{0_{4,L}}$ corresponds to the hypothesis that the operators projecting into $\mathcal{E}_L^{(1)}$ and $\mathcal{E}_L^{(2)}$ are identical, see section 4.4 for details.

Similar to above, a suitable test statistics is given by

$$D_{4,L} \stackrel{\text{def}}{=} \int\int \left\{ \sum_{r=1}^{L} \hat{\gamma}_r^{(1)}(t)\hat{\gamma}_r^{(1)}(s) - \sum_{r=1}^{L} \hat{\gamma}_r^{(2)}(t)\hat{\gamma}_r^{(2)}(s) \right\}^2 dtds.$$

The null hypothesis is rejected if $D_{4,L} \geq \Delta_{4,L;1-\alpha}$, where $\Delta_{4,L;1-\alpha}$ denotes the critical value of the distribution of

$$\begin{aligned}\Delta_{4,L} \stackrel{\text{def}}{=} \int\int \Bigg[&\sum_{r=1}^{L} \{\hat{\gamma}_r^{(1)}(t)\hat{\gamma}_r^{(1)}(s) - \gamma_r^{(1)}(t)\gamma_r^{(1)}(s)\} \\ &- \sum_{r=1}^{L} \{\hat{\gamma}_r^{(2)}(t)\hat{\gamma}_r^{(2)}(s) - \gamma_r^{(2)}(t)\gamma_r^{(2)}(s)\} \Bigg]^2 dtds.\end{aligned}$$

The distribution of $\Delta_{4,L}$ and hence its critical values are approximated by the bootstrap distribution of

$$\begin{aligned}\Delta_{4,L}^* \stackrel{\text{def}}{=} \int\int \Bigg[&\sum_{r=1}^{L} \{\hat{\gamma}_r^{(1)*}(t)\hat{\gamma}_r^{(1)*}(s) - \hat{\gamma}_r^{(1)}(t)\hat{\gamma}_r^{(1)}(s)\} \\ &- \sum_{r=1}^{L} \{\hat{\gamma}_r^{(2)*}(t)\hat{\gamma}_r^{(2)*}(s) - \hat{\gamma}_r^{(2)}(t)\hat{\gamma}_r^{(2)}(s)\} \Bigg]^2 dtds.\end{aligned}$$

It will be shown in Theorem 4.4 below, that under the null hypothesis as well as under the alternative, the distributions of $n\Delta_1, n\Delta_{2,r}, n\Delta_{3,r}, n\Delta_{4,L}$ converge to continuous limit distributions which can be consistently approximated by the bootstrap distributions of $n\Delta_1^*, n\Delta_{2,r}^*, n\Delta_{3,r}^*, n\Delta_{4,L}^*$.

4.3.1 Theoretical Results

Let $n = (n_1 + n_2)/2$. Assume that asymptotically $n_1 = n \cdot q_1$ and $n_2 = n \cdot q_2$ for some fixed proportions q_1 and q_2.

The following notation is used, $\mathcal{X}_1 = \{X_1^{(1)}, \ldots, X_{n_1}^{(1)}\}$ and $\mathcal{X}_2 = \{X_1^{(2)}, \ldots, X_{n_2}^{(2)}\}$, to denote the observed samples of random functions.

THEOREM 4.4 *Assume that* $\{X_1^{(1)}, \ldots, X_{n_1}^{(1)}\}$ *and* $\{X_1^{(2)}, \ldots, X_{n_2}^{(2)}\}$ *are two* independent *samples of random functions each of which satisfies the assumptions of the theorem 4.1.*

As $n \to \infty$ *we then obtain* $n\Delta_1 \xrightarrow{\mathcal{L}} F_1$, $n\Delta_{2,r} \xrightarrow{\mathcal{L}} F_{2,r}$, $n\Delta_{3,r} \xrightarrow{\mathcal{L}} F_{3,r}$, *and* $n\Delta_{4,L} \xrightarrow{\mathcal{L}} F_{4,L}$, *where* F_1, $F_{2,r}$, $F_{3,r}$, $F_{4,L}$ *are non-degenerated, continuous probability distributions. Furthermore, for any* $\delta > 0$

i)
$$|P(n\Delta_1 \geq \delta) - P(n\Delta_1^* \geq \delta |\ \mathcal{X}_1, \mathcal{X}_2)| = o_p(1)$$
as $n \to \infty$.

ii) If, furthermore, $\lambda_{r-1}^{(1)} > \lambda_r^{(1)} > \lambda_{r+1}^{(1)}$ *and* $\lambda_{r-1}^{(2)} > \lambda_r^{(2)} > \lambda_{r+1}^{(2)}$ *hold for some fixed* $r = 1, 2, \ldots,$ *then*
$$|P(n\Delta_{k,r} \geq \delta) - P\left(n\Delta_{k,r}^* \geq \delta |\ \mathcal{X}_1, \mathcal{X}_2\right)| = o_p(1), \quad k = 2, 3$$
as $n \to \infty$.

iii) If $\lambda_r^{(1)} > \lambda_{r+1}^{(1)}$ *and* $\lambda_r^{(2)} > \lambda_{r+1}^{(2)}$ *holds for all* $r = 1, \ldots, L$, *then*
$$|P(n\Delta_{4,L} \geq \delta) - P\left(n\Delta_{4,L}^* \geq \delta |\ \mathcal{X}_1, \mathcal{X}_2\right)| = o_p(1)$$
as $n \to \infty$.

Proof:
Proof and discussion of the structure of the distributions can be found in Benko et al. [2006b]. □

As discussed in the proceeding section, all curves in both samples are usually not directly observed, but have to be reconstructed from noisy observations according to (4.31). In this situation, the "true" empirical eigenvalues

and eigenfunctions have to be replaced by their discrete sample estimates. Bootstrap estimates are obtained by resampling the observations corresponding to the unknown curves $X_i^{(p)}$. As discussed in section 4.2.2, the validity of our test procedures is then based on the assumption that T is sufficiently large such that the additional estimation error is asymptotically negligible.

4.3.2 Simulation Study

In this paragraph we illustrate the finite behavior of the proposed test. We make use of the findings of the Example 4.2.7 and focus here on the test of common eigenfunctions. Looking at the figure 4.11 we observe that the error of the estimation of the eigenfunctions simulated by (4.38) is manifested by some shift of the estimated eigenfunctions. This motivates the basic simulation-setup (setup "a"), where the first sample is generated by the random combination of orthonormalized sine and cosine functions (Fourier functions) and the second sample is generated by the random combination of the same but shifted factor functions:

$$\begin{aligned} X_i^{(1)}(t_{ik}) &= \beta_{1i}^{(1)}\sqrt{2}\sin(2\pi t_{ik}) + \beta_{2i}^{(1)}\sqrt{2}\cos(2\pi t_{ik}) \\ X_i^{(2)}(t_{ik}) &= \beta_{1i}^{(2)}\sqrt{2}\sin\{2\pi(t_{ik}+\delta)\} + \beta_{2i}^{(2)}\sqrt{2}\cos\{2\pi(t_{ik}+\delta)\}. \end{aligned}$$

The factor loadings are i.i.d. random variables with $\beta_{1i}^{(p)} \sim N(0, \lambda_1^{(p)})$ and $\beta_{2i}^{(p)} \sim N(0, \lambda_2^{(p)})$. The functions are generated on the equidistant grid $t_{ik} = t_k = k/T,\ k = 1, \dots T = 100,\ i = 1, \dots, n = 70$. For the presentation of results in table 4.3, we use the following notation: "a) $\lambda_1^{(1)},\ \lambda_2^{(1)},\ \lambda_2^{(2)},\ \lambda_2^{(2)}$". The shift parameter δ is changing from 0 to 0.25 with the step 0.05. It should be mentioned that the shift $\delta = 0$ yields the simulation of level and setup with shift "$\delta = 0.25$" yields the simulation of the alternative, where the two factor functions are exchanged.

In the second setup (setup "b") the first factor functions are same and the second factor functions differ:

$$\begin{aligned} X_i^{(1)}(t_{ik}) &= \beta_{1i}^{(1)}\sqrt{2}\sin(2\pi t_{ik}) + \beta_{2i}^{(1)}\sqrt{2}\cos(2\pi t_{ik}) \\ X_i^{(2)}(t_{ik}) &= \beta_{1i}^{(2)}\sqrt{2}\sin\{2\pi(t_{ik}+\delta)\} + \beta_{2i}^{(2)}\sqrt{2}\sin\{4\pi(t_{ik}+\delta)\}. \end{aligned}$$

In the Table 4.3 we use the notation "b) $\lambda_1^{(1)},\ \lambda_2^{(1)},\ \lambda_2^{(2)},\ \lambda_2^{(2)}, D_r$". D_r means the test for the equality of the r-th eigenfunction. In the bootstrap tests we used 500 bootstrap replications. The critical level in this simulation is $\alpha = 0.1$. The number of simulations is 250.

We can interpret the Table 4.3 in the following way: in power simulations ($\delta \neq 0$) test behaves as expected: less powerful if the functions are "hardly

setup/shift	0	0.05	0.1	0.15	0.2	0.25
a) 10, 5, 8, 4	0.13	0.41	0.85	0.96	1	1
a) 4, 2, 2, 1	0.12	0.48	0.87	0.96	1	1
a) 2, 1,1.5, 2	0.14	0.372	0.704	0.872	0.92	0.9
b) 10, 5, 8, 4 D_1	0.10	0.44	0.86	0.95	1	1
b) 10, 5, 8, 4 D_2	1	1	1	1	1	1

Table 4.3: The results of the simulations for $\alpha = 0.1$, $n = 70$, $T = 100$, number of simulations is 250.

distinguishable" (small shift, small difference in eigenvalues). The level approximation seems to be less precise if the difference in the eingenvalues $(\lambda_1^{(p)} - \lambda_2^{(p)})$ becomes smaller, this can be explained by relative small sample-size n, small number of bootstrap-replications and increasing estimation-error as argued in the Theorem 2, assertion (iii).

In comparison to our general setup (4.31) we used an equidistant and common design for all functions. This simplification is necessary, it simplifies and speeds-up the simulations, in particular using general random and observation-specific design makes the simulation computationally untractable.

Secondly, we omitted the additional observation error, this corresponds to the standard assumptions in the functional principal components theory. As argued in the section 4.2.2 the inference based on the directly observed functions and estimated functions X_i is first order equivalent under mild conditions implied by Theorems 4.3 and 4.4. In order to illustrate this theoretical result in the simulation we used the following setup:

$$\begin{aligned} X_i^{(1)}(t_{ik}) &= \beta_{1i}^{(1)}\sqrt{2}\sin(2\pi t_{ik}) + \beta_{2i}^{(1)}\sqrt{2}\cos(2\pi t_{ik}) + \varepsilon_{ik}^{(1)} \\ X_i^{(2)}(t_{ik}) &= \beta_{1i}^{(2)}\sqrt{2}\sin\{2\pi(t_{ik}+\delta)\} + \beta_{2i}^{(2)}\sqrt{2}\cos\{2\pi(t_{ik}+\delta)\} + \varepsilon_{ik}^{(2)}, \end{aligned}$$

where $\varepsilon_{ik}^{(p)} \sim N(0, 0.25)$, $p = 1, 2$. All other parameters remain same as in the simulation setup "a". Using this setup we recalculate the simulation presented in the second "line" of the table 4.3, for estimation of the functions $X_i^{(p)}$, $p = 1, 2$ we used Nadaraya-Watson estimation with Epanechnikov kernel and bandwidth $b = 0.05$. We run the simulations with various bandwidths, the choice of the bandwidth doesn't have strong influence on results except by oversmoothing (large bandwidths). The results are printed in the Table 4.4. As we can see the difference of the simulation results using estimated functions are not significantly different in comparison to the results printed in the second line of the Table 4.3 – directly observed functional

setup/shift	0	0.05	0.1	0.15	0.2	0.25
a)10,5,8,4	0.09	0.35	0.64	0.92	0.94	0.97

Table 4.4: The results of the simulation for $\alpha = 0.1$, $n = 70$, $T = 100$ with additional error in observation.

values.

The last limitation of this simulation study is the choice of particular alternative. A more general setup of this simulation study might be based on the following model: $X_i^{(1)}(t) = \beta_{1i}^{(1)}\gamma_1^{(1)}(t) + \beta_{2i}^{(1)}\gamma_2^{(1)}(t)$, $X_i^{(2)}(t) = \beta_{1i}^{(2)}\gamma_1^{(2)}(t) + \beta_{2i}^{(2)}\gamma_2^{(2)}(t)$ where $\gamma_1^{(1)}, \gamma_1^{(2)}, \gamma_2^{(1)}$ and g are mutually orthogonal functions on $L^2[0,1]$ and $\gamma_2^{(2)} = (1+\upsilon^2)^{-1/2}\{\gamma_2^{(1)} + \upsilon g\}$. Basically we create the alternative by the contamination of one of the "eigenfunctions" (in our case the second one) in the direction g and ensure $||\gamma_2^{(2)}|| = 1$. The amount of the contamination is controlled by the parameter υ. Note that the exact squared integral difference $||\gamma_2^{(1)} - \gamma_2^{(2)}||^2$ does not depend on function g. Thus in the "functional sense" particular "direction of the alternative hypothesis" represented by the function g has no impact on the power of the test. However, since we are using nonparametric estimation technique, we might expect that rough (highly fluctuating) functions g will yield higher error of estimation and hence decrease the precision (and power) of the test. Finally, higher number of factor functions (L) in simulation may cause less precise approximation of critical values and more bootstrap replications and larger sample-size may be needed. This can also be expected from the Theorem 2 in section 4.2.2 – the variance of the estimated eigenfunctions depends on all eigenfunctions corresponding to non-zero eigenvalues.

4.4 Further Remarks

Dual and Functional Basis Approach
The dual approach presented in the section 4.2.6 can be seen as a special case of the functional basis approach, where the basis functions are the functions $X_i, i = 1 \ldots, n$. The coefficient matrix $\mathbf{C}$ is a n-dimensional identity matrix and matrix M corresponds to the cross-term matrix $\mathbf{W}$ in section 4.2.3, where the whole information is stored. Otherwise, the whole notation introduced in the section 4.2.3 remains valid.
Common Eigefunctions Hypothesis
Please note that the eigenfunctions γ_r are identified only up to sign and the eigenfunctions with correct sign needs to be taken so that the theorem 4.4 can be applied. This needs be considered by the construction of the statistics $D_{2,r}$ and its bootstrap variant $\Delta^*_{2,r}$:

$$\begin{aligned} D_{2,r}(\gamma_r^{(1)}, \gamma_r^{(2)}) &\stackrel{\text{def}}{=} \|\hat{\gamma}_r^{(1)} - \hat{\gamma}_r^{(2)}\|^2, \\ \Delta^*_{2,r}(\gamma_r^{(1)}, \gamma_r^{(1)*}, \gamma_r^{(2)}, \gamma_r^{(2)*}) &\stackrel{\text{def}}{=} \|\hat{\gamma}_r^{(1)*} - \hat{\gamma}_r^{(1)} - (\hat{\gamma}_r^{(2)*} - \hat{\gamma}_r^{(2)})\|^2. \end{aligned}$$

Benko et al. [2006b] proposed the following corrections:
$D_{2,r} = \min(D_{2,r}(\gamma_r^{(1)}, \gamma_r^{(2)}), D_{2,r}(\gamma_r^{(1)}, -\gamma_r^{(2)}))$ and
$\Delta^*_{2,r} = \min(\Delta^*_{2,r}((-1)^j\gamma_r^{(1)}, (-1)^k\gamma_r^{(1)*}, (-1)^l\gamma_r^{(2)}, (-1)^p\gamma_r^{(2)*}), j, k, l, p = 0, 1)$.
Common Eigenspaces Hypothesis
As already argued the hypothesis of common eigenspaces corresponds naturally to the equality of the projection operators projecting into these spaces. More detailed, for any v denote its projection $v_{\mathcal{E}_L^{(1)}}$ into the eigenspace: $\text{span}\{\gamma_1^{(1)}, \ldots, \gamma_L^{(1)}\}$, then:

$$v_{\mathcal{E}_L^{(1)}} = \sum_{r=1}^{L} < v, \gamma_r^{(1)} > \gamma_r^{(1)}$$

similarly for the projection $v_{\mathcal{E}_L^{(2)}}$ into $\text{span}\{\gamma_1^{(2)}, \ldots, \gamma_L^{(2)}\}$:

$$v_{\mathcal{E}_L^{(2)}} = \sum_{r=1}^{L} < v, \gamma_r^{(2)} > \gamma_r^{(2)}.$$

If the two eigenspaces are equal ($v_{\mathcal{E}_L^{(1)}} = v_{\mathcal{E}_L^{(2)}}$) then

$$\begin{aligned}\sum_{r=1}^{L} < v, \gamma_r^{(1)} > \gamma_r^{(1)(t)} &= \sum_{r=1}^{L} (\int v(s)\gamma_r^{(1)}(s)ds)\gamma_r^{(1)(t)} \\ &= \sum_{r=1}^{L} (\int v(s)\gamma_r^{(1)}(s)ds)\gamma_r^{(1)(t)} \\ &= \sum_{r=1}^{L} < v, \gamma_r^{(2)} > \gamma_r^{(2)}(t)\end{aligned}$$

for all t. Since this holds for all functions v this yield

$$\sum_{r=1}^{L} \gamma_r^{(1)}(t)\gamma_r^{(1)}(s) = \sum_{r=1}^{L} \gamma_r^{(2)}(t)\gamma_r^{(2)}(s) \quad \text{for all } t, s \in [0,1]. \tag{4.40}$$

Calculation of $\Delta_{4,L}$.
In the implementation the double integral in $\Delta_{4,L}$ does not need to be evaluated directly since:

$$\begin{aligned}D_{4,L} &\stackrel{\text{def}}{=} \int\int \left\{ \sum_{r=1}^{L} \hat{\gamma}_r^{(1)}(t)\hat{\gamma}_r^{(1)}(s) - \sum_{r=1}^{L} \hat{\gamma}_r^{(2)}(t)\hat{\gamma}_r^{(2)}(s) \right\}^2 dtds \\ &= \int\int \left\{ \sum_{r=1}^{L} \hat{\gamma}_r^{(1)}(t)\hat{\gamma}_r^{(1)}(s) \right\}^2 + \left\{ \sum_{r=1}^{L} \hat{\gamma}_r^{(2)}(t)\hat{\gamma}_r^{(2)}(s) \right\}^2 \\ &\quad -2 \left\{ \sum_{r=1}^{L} \hat{\gamma}_r^{(1)}(t)\hat{\gamma}_r^{(1)}(s) \sum_{r=1}^{L} \hat{\gamma}_r^{(2)}(t)\hat{\gamma}_r^{(2)}(s) \right\} dtds \\ &= 2L - 2 \left\{ \sum_{r=1}^{L} \sum_{q=1}^{L} \langle \hat{\gamma}_q^{(1)}, \hat{\gamma}_r^{(2)} \rangle^2 \right\}\end{aligned}$$

and hence only simple one-dimensional integrals need to be calculated. In the application these integrals can be approximated by the Riemann-sums calculated on the sufficiently fine grid – this grid can be arbitrary fine since the final estimators (4.35) can be calculated on any arbitrary fine grid.

A simple example – case of $H_{0_{4,2}}$, may contribute to better understanding: $L = 2,\ H_{0_{4,2}}$,: $\gamma_1^{(2)} = c_1\gamma_1^{(1)} + c_2\gamma_2^{(1)}$,$\gamma_2^{(2)} = d_1\gamma_1^{(1)} + d_2\gamma_2^{(1)}$, where $c_1^2 + c_2^2 = 1$,

$d_1^2 + d_2^2 = 1$ and $c_1 d_1 + c_2 d_2 = 0$, due to orthonormality of $\gamma_1^{(2)}, \gamma_2^{(2)}$. Then

$$\begin{aligned} D_{4,2} &= 4 - 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)},\gamma_r^{(2)}\rangle^2\right\} \\ &= 4 - 2\left\{\langle\gamma_1^{(1)},\gamma_1^{(2)}\rangle^2 + \langle\gamma_1^{(1)},\gamma_2^{(2)}\rangle^2 + \langle\gamma_2^{(1)},\gamma_1^{(2)}\rangle^2 + \langle\gamma_2^{(1)},\gamma_2^{(2)}\rangle^2\right\} \\ &= 4 - 2\left\{\langle\gamma_1^{(1)}, c_1\gamma_1^{(1)} + c_2\gamma_2^{(1)}\rangle^2 + \langle\gamma_1^{(1)}, d_1\gamma_1^{(1)} + d_2\gamma_2^{(1)}\rangle^2\right. \\ &\quad \left. + \langle\gamma_2^{(1)}, c_1\gamma_1^{(1)} + c_2\gamma_2^{(1)} + \langle\gamma_1^{(1)}, d_1\gamma_2^{(1)} + d_2\gamma_2^{(1)}\rangle^2\right\} \\ &= 4 - 2\{c_1^2 + d_1^2 + c_2^2 + d_2^2\} = 0 \end{aligned}$$

It might be also interesting to compare the $D_{4,1}$ and $D_{2,1}$: $L = 1$: $D_{4,1} = 2 - 2\langle\hat{\gamma}_1^{(1)}, \hat{\gamma}_1^{(2)}\rangle^2$ (compare with $D_{2,1} = 2 - 2\langle\hat{\gamma}_r^{(1)}, \hat{\gamma}_r^{(2)}\rangle$).
While $D_{4,1}$ is invariant w.r.t. sign of the eigenfunctions, the $D_{2,1}$ is not. Since these two statistics can be considered as equivalent an alternative statistics for testing the hypothesis $D_{2,r}$ might be also based on

$$\tilde{D}_{4,r} = 2 - 2\langle\hat{\gamma}_r^{(1)}, \hat{\gamma}_r^{(2)}\rangle^2.$$

Calculation of $\Delta_{4,L}^*$ The bootstrap statistic $\Delta_{4,L}^*$ can be calculated using same arguments by:

$$\begin{aligned} \Delta_{4,L}^* &= \int\int\left[\sum_{r=1}^{L}\{\gamma_r^{(1)*}(t)\gamma_r^{(1)*}(s) - \gamma_r^{(1)}(t)\gamma_r^{(1)}(s)\}\right. \\ &\quad \left. - \sum_{r=1}^{L}\{\gamma_r^{(2)*}(t)\gamma_r^{(2)*}(s) - \gamma_r^{(2)}(t)\gamma_r^{(2)}(s)\}\right]^2 dtds \\ &= \int\int\left[\sum_{r=1}^{L}\{\gamma_r^{(1)*}(t)\gamma_r^{(1)*}(s) - \gamma_r^{(1)}(t)\gamma_r^{(1)}(s)\}\right]^2 + \\ &\quad \left[\sum_{r=1}^{L}\{\gamma_r^{(2)*}(t)\gamma_r^{(2)*}(s) - \gamma_r^{(2)}(t)\gamma_r^{(2)}(s)\}\right]^2 \\ &\quad -2\sum_{r=1}^{L}\{\gamma_r^{(1)*}(t)\gamma_r^{(1)*}(s) - \gamma_r^{(1)}(t)\gamma_r^{(1)}(s)\} \\ &\quad \sum_{r=1}^{L}\{\gamma_r^{(2)*}(t)\gamma_r^{(2)*}(s) - \gamma_r^{(2)}(t)\gamma_r^{(2)}(s)\}dsdt \end{aligned}$$

$$
\begin{aligned}
= \quad & 2L - 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)},\gamma_r^{(1)*}\rangle^2\right\} + 2L - 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(2)},\gamma_r^{(2)*}\rangle^2\right\} \\
& -2\int\int\{\sum_{r=1}^{L}\sum_{q=1}^{L}\{\gamma_r^{(1)*}(t)\gamma_r^{(1)*}(s) - \gamma_r^{(1)}(t)\gamma_r^{(1)}(s)\}. \\
& .\{\gamma_q^{(2)*}(t)\gamma_q^{(2)*}(s) - \gamma_q^{(2)}(t)\gamma_q^{(2)}(s)\}dsdt
\end{aligned}
$$

$$
\begin{aligned}
= \quad & 2L - 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)},\gamma_r^{(1)*}\rangle^2\right\} + 2L - 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(2)},\gamma_r^{(2)*}\rangle^2\right\} \\
& -2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)*},\gamma_r^{(2)*}\rangle^2\right\} + 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)*},\gamma_r^{(2)}\rangle^2\right\} \\
& +2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)},\gamma_r^{(2)*}\rangle^2\right\} - 2\left\{\sum_{r=1}^{L}\sum_{q=1}^{L}\langle\gamma_q^{(1)},\gamma_r^{(2)}\rangle^2\right\}
\end{aligned}
$$

Again only one-dimensional integrals (scalar products) need to be evaluated.

Bibliographic Notes

Dual Approach: The estimation proposed in the section is an adaptation of the method proposed by Kneip and Utikal [2001] for construction of the factor model for the sample of densities: $f_i, i = 1, \ldots, n$, i.e. $X_i = f_i$. In order to avoid complications in the notation, assume that this sample is centered. Using similar arguments, the FPCA on this sample is performed on the scalar product matrix:

$$M_{ij} = \langle f_i, f_j\rangle. \tag{4.41}$$

Assuming the standard density estimation problem for each i, as described in section 3.4, and denoting the number of observed realisations for $i - th$ problem by T_i, Kneip and Utikal [2001] proposed to plug-in a KDE estimate (3.24):

$$\hat{f}_{i,h}(t) = \frac{1}{Th}\sum_{j=1}^{T_i} K\left(\frac{t - t_{ij}}{h}\right), \tag{4.42}$$

where t_{ij} is the jth of the sample generated by the density f_i) in (4.41). The linear structure of (4.42) enables to construct the estimate of (4.41) by

bias-corrected version of:

$$\tilde{M}_{ij} = \frac{1}{T_i T_j h^2} \sum_{l=1}^{T_i} \sum_{k=1}^{T_j} \int K\left(\frac{t - t_{il}}{h}\right) K\left(\frac{t - t_{jk}}{h}\right) dt$$

the remaining part of the estimation procedure corresponds to the procedure in the regression case described in section 4.2.6 except that in the formula (4.35) not a linear smoother but a KDE. In general different estimate than that used in (4.42) is used to estimate $\hat{X}_i = \hat{f}_i$.

This approach is of particular importance, since if analyzing the sample of densities a KDE approach is straighforward whereas a FPCA based on the functional basis expansion might be considered as complicated in case of densities.

Using this methodology Kneip and Utikal [2001] proposed a dimension test (choice of L in a model corresponding to the truncated KL-expansion 4.11) based on the generation of the bootstrap samples from (4.11).

Common principal components: As already mentioned common principal component methodology has been proposed in the multivariate statistics, see monograph on CPC, Flury [1988]. Flury's methodology introduce CPC in spirit of testing the similarities in the covariance matrices. The testing procedure relies on maximum likelihood principle and assumes normality. The estimation procedure is based on the numerical procedure – simultaneous diagonalization of covariance matrices – Flury-Gautschi algorithm. Generalizations to FPCA have been discussed in Ferraty et al. [2006] and Viguier-Pla [2004]. Their methodology is based on the asymptotic arguments direct application seems to be difficult and are in practice the application has to be based on the discretization or projection of the functional objects to the finite dimensional space.

In the empirical finance CPC methodology has been first applied (in multivariate setup) to forward rate models by Alexander [2002] and to IV analysis by Fengler et al. [2003]. Flury-Gautschi algorithm have been used for estimating CPC in the FPCA in an applied IV analysis by Benko and Härdle [2005]. Application of the methods proposed in this chapter to the IVs analysis is discussed in the next chapter.

4.5 Implied Volatility Analysis

In this section we present an application of the method discussed in previous sections to the implied volatilities of european options on the German stock index (ODAX). The concept, interpretation and application of the (Black-Scholes) IV, studied here, is deeply discussed in the chapter 2, the computational and practical issues connected with the calculation and static estimation of the IV is studied in the chapter 3. This section goes beyond the scope of the static modeling (estimation of the IV function or IVS for given time point) and discusses the dynamic models for time-development of the IVS.

Fengler et al. [2003] studied the dynamics of the IV via PCA on discretized IV functions for different maturity groups and tested the Common Principal Components (CPC) hypotheses (equality of eigenvectors and eigenspaces for different groups). Their method rely on the CPC methodology introduced by Flury [1988] which is based on maximum likelihood estimation under the assumption of multivariate normality. The main aim of this application is to verify their results in a functional sense. Doing so, we overcome two basic weaknesses of their approach. Firstly, the factor model proposed by Fengler et al. [2003] is just performed on a sparse design of moneyness. However, in practice, e.g. in Monte-Carlo pricing methods evaluation on a fine grid is needed. Using the functional PCA approach we may overcome this difficulty and evaluate the factor model on an arbitrary fine grid. A second difficulty of the procedure proposed by Fengler et al. [2003] comes from the data design – on the exchange we cannot observe the option with desired maturity on each day and we need to estimate them from the IV-functions with maturities observed on the particular day. Consequently the two-dimensional Nadaraya-Watson estimator proposed by Fengler et al. [2003] results essentially in the (weighted) average of the IVs (with closest maturities) observed on particular day, which may affect the test of the common eigenfunction hypothesis. We use the linear interpolation scheme in the *total variance* $\sigma^2_{TOT,i}(\kappa,\tau) \stackrel{\text{def}}{=} (\sigma^\tau_i(\kappa))^2\tau$, in order to recover the IV functions with fixed maturity (on day i). This interpolation scheme is based on the arbitrage arguments originally proposed by Kahalé [2004] for zero-divident and zero-interest rate case and generalized for deterministic interest rate by Fengler [2005a], see also section 2.4.2.

More precisely, having IVs with maturities observed on a particular day i: $\tilde\sigma_i^{\tau_{j_i}}(\kappa)$, $j_i = 1,\ldots,p_{\tau_i}$, we calculate the corresponding total variance $\tilde\sigma_{TOT,i}(\kappa,\tau_{j_i})$. From these total variances we linearly interpolate the total variance with the desired maturity from the nearest maturities observed on day i. The total variance can easily be transformed to corresponding IV

$\tilde{\sigma}_i^\tau(\kappa)$. As the last step we calculate the log-returns

$$\triangle \log \tilde{\sigma}_i^\tau(\kappa) \stackrel{\text{def}}{=} \log \tilde{\sigma}_{i+1}^\tau(\kappa) - \log \tilde{\sigma}_i^\tau(\kappa).$$

The log-IV-returns are observed for each maturity τ on a discrete grid κ_{ik}^τ. We assume that observed log-IV-return $\triangle \log \tilde{\sigma}_i^\tau(\kappa_{ik}^\tau)$ consists of true log-return of the IV function denoted by $\triangle \log \sigma_i^\tau(\kappa_{ik}^\tau)$ and possibly of some additional error ε_{ik}^τ. By setting

$$\begin{aligned} Y_{ik}^\tau &:= \triangle \log \tilde{\sigma}_i^\tau(\kappa_{ik}^\tau), \\ X_i^\tau(\kappa) &:= \triangle \log \sigma_i^\tau(\kappa) \end{aligned}$$

we obtain analogue of the model (4.31) with the argument κ:

$$Y_{ik}^\tau = X_i^\tau(\kappa_{ik}) + \varepsilon_{ik}^\tau, \ i = 1, \ldots, n_\tau. \tag{4.43}$$

In order to simplify the notation and make the connection with the theoretical part clear we will use the notation in form of (4.43).

For our analysis we use a recent data set containing the daily data from 1st January 2004 to 15th June 2004 taken from the German-Swiss exchange EUREX. The violations of the arbitrage-free assumptions were corrected using procedure proposed by Fengler [2005a]. Similar to Fengler et al. [2003] we excluded options with maturity smaller than 10 days, these option-prices are known to be very noisy, partially because of a special and arbitrary setup in the pricing systems of the dealers. Using the interpolation scheme described above we calculate the log-IV-returns for two maturity groups $\tau = 0.12$ (measured in years), we denote it as "1M" group and $\tau = 0.36$ ("3M" group) and denote them by

$$\begin{aligned} &Y_{ik}^{1M}, \ k = 1, \ldots, K_i^{1M}, \\ &Y_{ik}^{3M}, \ k = 1, \ldots, K_i^{3M}. \end{aligned}$$

Since we ensured that for each i, the interpolation procedure does not use data with same maturity for both groups, this procedure has no impact on the independence of both samples. The underlying models, based on the truncated version of (4.12) are:

$$X_i^{1M}(\kappa) = \bar{X}_i^{1M}(\kappa) + \sum_{r=1}^{L_{1M}} \hat{\beta}_{ri}^{1M} \widehat{\gamma}_r^{\,1M}(\kappa), i = 1, \ldots, n_{1M} \tag{4.44}$$

$$X_i^{3M}(\kappa) = \bar{X}_i^{3M}(\kappa) + \sum_{r=1}^{L_{3M}} \hat{\beta}_{ri}^{3M} \widehat{\gamma}_r^{\,3M}(\kappa), i = 1, \ldots, n_{3M}. \tag{4.45}$$

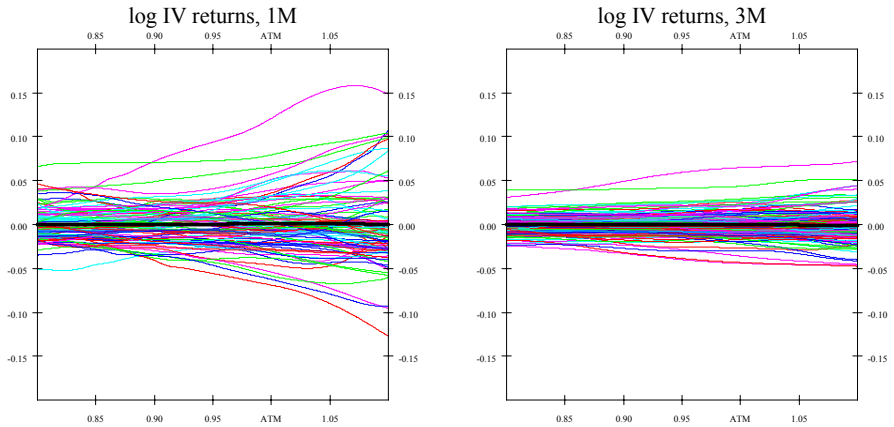

Figure 4.12: Nadaraya-Watson estimator of the log-IV-returns for maturity 1M in left figure and 3M in right figure. The bold line is the sample mean of the corresponding group.

Model (4.44) and (4.45) can serve e.g. in a Monte Carlo pricing tool in the risk management for pricing exotic options where the whole path of implied volatilities is needed to determine the price. Estimating the factor functions in (4.44) and (4.45) by eigenfunctions displayed in figure 4.13 we only need to fit the (estimated) factor loadings $\hat{\beta}_{ji}^{1M}$ and $\hat{\beta}_{ji}^{3M}$. The pillar of the model is the dimension reduction. Keeping the factor function fixed for a certain time period we need to analyze (two) multivariate random processes of the factor loadings. For the purposes of this paper we will concentrate on comparing the factors of the models (4.44) and (4.45) and the technical details of the analysis of the factor loading will not be discussed here, we refer to Fengler et al. [2003], who proposed to fit the factor loadings by centered normal distributions with diagonal variance matrix containing the corresponding eigenvalues. For a deeper discussion of the fitting of factor loadings using a more sophisticated approach, basically based on (possibly multivariate) GARCH models, see Fengler [2005b].

From our data set we obtained 88 functional observations for the 1M group (n_{1M}) and 125 observations for the 3M group (n_{3M}). We will estimate the model on the interval for futures moneyness $\kappa \in [0.8, 1.1]$. In comparison to Fengler et al. [2003] we may estimate the models (4.44) and (4.45) on arbitrary fine grid (we used an equidistant grid of 500 points on the interval $[0.8, 1.1]$). For illustration, the Nadaraya-Watson estimator of resulting log-returns is plotted in figure 4.12. The smoothing parameters have been

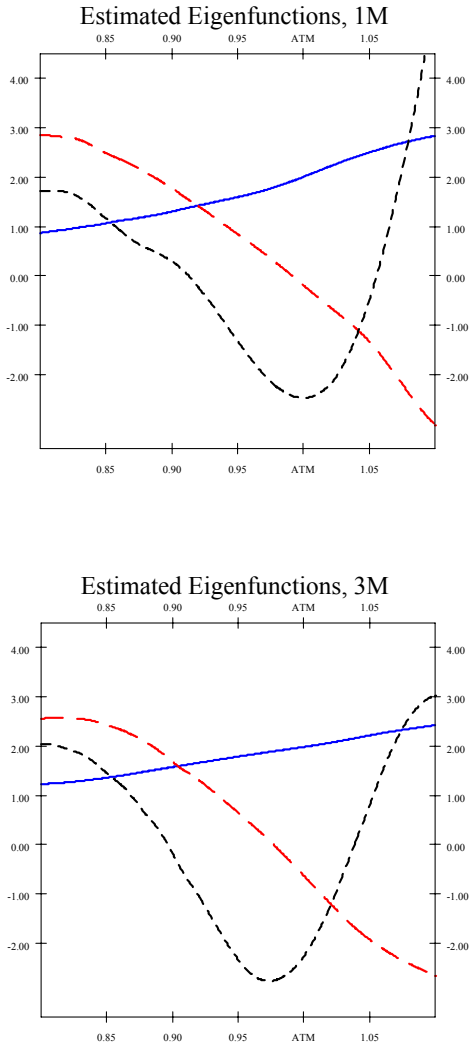

Figure 4.13: Estimated eigenfunctions for 1M group in the left plot and 3M group in the right plot, blue solid – first function, red dashed – second function, black finely dashed – third function.

chosen in accordance with the requirements in section 4.2.2. As argued in section 4.2.2, we should use small smoothing parameters in order to avoid a possible bias in the estimated eigenfunctions. Thus we use for each i essentially the smallest bandwidth b_i that guarantees that estimator $\hat{X}_i$ is defined on the whole support $[0.8, 1.1]$.

Using the procedures described in section 4.2.2 we first estimate the eigenfunctions of the both maturity groups. The estimated eigenfunctions are plotted in figure 4.13. The structure of the eigenfunctions is in accordance with other empirical studies on IV-surfaces, for a deeper discussion and economic interpretation see for example Fengler et al. [2005] or Fengler et al. [2003].

Clearly, the ratio of the variance explained by the k-th factor function is given by the quantity

$$\hat{\nu}_k^{1M} = \hat{\lambda}_k^{1M} / \sum_{j=1}^{n_{1M}} \hat{\lambda}_j^{1M},$$

for the 1M group, correspondingly $\hat{\nu}_k^{3M}$ for the 3M group. In Table 4.5 we list the contributions of the factor functions. Looking at the Table 4.5 we can see, that the 4-th factor functions explain less than 1% of the variation, this number was the "threshold" for the choice of the L_{1M} and L_{2M}.

	var. explained 1M	var. explained 3M
$\hat{\gamma}_1^\tau$	89.9%	93.0%
$\hat{\gamma}_2^\tau$	7.7%	4.2%
$\hat{\gamma}_3^\tau$	1.7%	1.0%
$\hat{\gamma}_4^\tau$	0.6%	0.4%

Table 4.5: Variance explained by the eigenfunctions.

We can observe, see figure 4.13, that the factor functions for both groups are similar. Thus, in the next step we use the bootstrap test for testing the equality of the factor functions. We use 2000 bootstrap replications. The test of equality of the eigenfunctions was rejected for the first eigenfunction for the analyzed time period (January 2004 – June 2004) at a significance level $\alpha = 0.05$ (P-value 0.01). We may conclude that the (first) factor functions are not exactly same in the factor model for both maturity groups. However from a practical point of view we are more interested in the checking the appropriateness of the whole models for fixed number of factors: $L = 2$ or $L = 3$ in (4.44) and (4.45), this turns into testing the equality of eigenspaces. Thus, in the next step we test with the same setup (2000 bootstrap replications) the hypotheses that first two and first three eigenfunctions span the

same eigenspaces $\mathcal{E}_L^{1M}$ and $\mathcal{E}_L^{3M}$. Both hypotheses $L = 2$ and $L = 3$ are not rejected at the significance level $\alpha = 0.05$ (P-value 0.61 for $L = 2$ and 0.09 for $L = 3$). Summarizing, even in the functional sense we have no significant reason to reject the hypothesis of common eigenspaces for these two maturity groups. Using this hypothesis the factors governing the movement of the returns of IV surface are invariant to time to maturity, just their relative importance can change. This leads to the common factor model:

$$X_i^\tau(\kappa) = \bar{X}^\tau(\kappa) + \sum_{r=1}^{L_\tau} \hat{\beta}_{ri}^\tau \widehat{\gamma}_r(\kappa), i = 1, \ldots, n_\tau,\ \tau = 1M, 3M.$$

Where $\gamma_r := \gamma_r^{1M} = \gamma_r^{3M}$. Besides the contribution to the understanding the structure of the IV function dynamics, in the sense of dimension reduction, using the common factor model we reduce the number of functional factors by half comparing to models (4.44) and (4.45). Furthermore, from the technical point of view, we also obtain an additional dimension reduction and higher estimation precision, since under this hypothesis we may estimate the eigenfunctions from the (individually centered) pooled sample $X_i(\kappa)^{1M}, i = 1, \ldots, n_{1M},\ X_i^{3M}(\kappa), i = 1, \ldots, n_{3M}\}$. The main improvement in comparison to the multivariate study by Fengler et al. [2003] is that our test performed in the functional sense, does not depend on particular discretization and our factor model can be evaluated on an arbitrary fine grid.

4.5.1 Further Remarks on IVs-Factor Analysis

Factor models in IV analysis are popular choice in empirical-finance literature. IVs calculated from the DAX options have been analyzed using PCA methodology (besides already mentioned work) by Fengler et al. [2002] (focused on the term-structure of the IVs), Cont and da Fonseca [2002], recently revisited by Detlefsen and Härdle [2006], who performed the IV analysis on extraordinary long time-to-maturity interval (up to 4 years) among others. Last two papers are proposing factor models for the whole IVS, using notation of section 4.5 following model is assumed:

$$X_i(\kappa, \tau) = \bar{X}(\kappa, \tau) + \sum_{r=1}^{L} \beta_{ri} \gamma_r(\kappa, \tau), i = 1, \ldots, n, \tag{4.46}$$

where $X_i(\kappa, \tau)$ are two dimensional IVS, $\bar{X}(\kappa, \tau)$ is the mean surface for the given period, β_{ri} (scalar) factor loadings and $\gamma_r(\kappa, \tau)$ two dimensional factor functions (surfaces).

The estimation strategy is similar to that used in already mentioned work of Fengler et al. [2003]:

1. estimate $X_i(\kappa, \tau)$ on daily basis, by nonparametric techniques (NW estimate has been employed) on a dense grid

2. perform PCA on the resulting (discrete) grid of functional values

Cont and da Fonseca [2002] extract 3 factor model, for S&P 500 Options (Standard&Poors) IVS on support $\kappa \in [0.5, 1.5]$ and $\tau \in [0, 1.5]$, first factor with the level interpretation (explaining 94% of the daily sample variation), second slope (explaining 3% of the sample variation) and third factor function the twist effect (0.8% of the variation), and similar results for FTSE 100 (Financial Times Stock Exchange index) options (level-factor 96% of the sample variation, slope-factor 2% of the sample variation and curvature-factor 0.8% of the sample variation), the shape of the factor functions obtained by Cont and da Fonseca [2002] are "homogeneous" in τ direction and seems to support the common principal components approach.

Detlefsen and Härdle [2006] identify 5 factor functions for ODAX IVS on support $\kappa \in [0.75, 1.25]$ and $\tau \in [0, 4]$. The factor functions are displayed in figure B.8 in the appendix. Results of Detlefsen and Härdle [2006] suggests that the option with long maturity require the factor model with much more complicated structure. At the present time it seems to be hard to investigate the term structure of the IV for longer maturities by the common FPCA approach proposed here, since only few observation are available for longer maturities. However following the current trend of trading options with longer maturities, see also section 3.6.1, it might be an interesting topic of the further research to investigate the common FPCA model for longer maturities.

It should be noted that, by estimating the surfaces for each day, one needs to essentially ensure that estimate is defined for whole analyzed support. This leads to either small analyzed support (close to ATM) and short time-to-maturities or to possible oversmoothing of out-coming factor functions.

An interesting approach, that seems to overcame this problem has been proposed by Fengler et al. [2005]. This approach combines both steps (estimate X_i, estimate factor model, e.g. by PCA) into one by assuming

$$Y_i(\kappa_{ji}, \tau_{ji}) = \gamma_0(\kappa_{ji}, \tau_{ji}) + \sum_{r=1}^{L} \beta_{ri}\gamma_r(\kappa_{ji}, \tau_{ji}) + \varepsilon_{ji}, \tag{4.47}$$

for $j = 1, \ldots, T_i, i = 1, \ldots, n$.

The estimates $\hat{\beta}_{ri}$ and $\hat{\gamma}_r$ of β_{ri} and γ_r respectively are given as minimizers

of the following least squares criterion:

$$\sum_{i=1}^{n}\sum_{j=1}^{T_i}\int\left\{Y_i(\kappa_{ji},\tau_{ji})-\sum_{r=0}^{L}\hat{\beta}_{ri}\hat{\gamma}_r(\kappa_{ji},\tau_{ji})\right\}^2\mathcal{K}_h[(\kappa,\tau)-(\kappa_{ji},\tau_{ji})]d(\kappa,\tau) \tag{4.48}$$

with $\hat{\beta}_{i,0}\stackrel{\text{def}}{=}1$. Optimization problem (4.48) is solved by iterative solution of the following equations:

$$\sum_{i=1}^{n}T_i\hat{\beta}_{i,r'}\hat{q}_i(\kappa,\tau) = \sum_{i=1}^{n}\sum_{r=0}^{L}\hat{\beta}_{i,r'}\hat{\beta}_{i,r}\hat{p}_i(\kappa,\tau)\hat{\gamma}_r(\kappa,\tau), \tag{4.49}$$

$$\int\hat{q}_i(\kappa,\tau)\hat{\gamma}_{r'}(\kappa,\tau) = \sum_{r=0}^{L}\hat{\beta}_{i,r}\int\hat{p}_i(\kappa,\tau)\hat{\gamma}_{r'}(\kappa,\tau)\hat{\gamma}_r(\kappa,\tau)d(\kappa,\tau). \tag{4.50}$$

with $\hat{p}_i(\kappa,\tau)=T_i^{-1}\sum_{j=1}^{T_i}\mathcal{K}_h[(\kappa,\tau)-(\kappa_{ji},\tau_{ji})]$ and $\hat{q}_i(\kappa,\tau)=T_i^{-1}\sum_{j=1}^{T_i}\mathcal{K}_h[(\kappa,\tau)-(\kappa_{ji},\tau_{ji})]Y_i(\kappa_{ji},\tau_{ji})$. Since the estimates $\hat{\beta}_{ri}$ and $\hat{\gamma}_r$ of β_{ri} and γ_r are not uniquely defined by the solution of (4.48) these are, in addition, normalized essentially by an L^2 orthogonalization procedure. Basically, the method is motivated by some back-fitting arguments – cyclical fitting of factor functions $\gamma_r(\kappa_{ji},\tau_{ji})$ based on the equations (4.50) and (4.50). The estimation procedure can be seen as an adaptation of the well known back-fitting algorithm used in the additive models, see Hastie et al. [2002]. This approach – refereed to as Dynamic Semiparametric Factor Model (DSFM) seems to mild the possible oversmoothing problems caused by the sparse design of the IVs observed on particular day, since it is using all data (not only the the data from a single day) for estimation of the individual functional objects γ_r and only the scalar factor loadings are estimated from the "daily" data. On the other hand, as known from standard additive models the inference for the DSFM models are difficult to handle. Unlike a factor model based directly on the (functional) principal components, as discussed in the previous part of this chapter, the DSFM models for different L are not nested and its estimation is connected with high computational costs.

Appendix A

FDA Library

The library for functional data analysis, designed for the statistical environment XploRe is presented in this chapter. The library consists of XploRe-macros (Quantlets) summarized into the XploRe library (quantlib) `fda`. The library `fda` is a joint work with J. Ulbricht, see also Ulbricht [2004] and Benko [2004]. For similar packages for R-project, software S and Matlab see Ramsay [2003], see also section A.1.4 for further comments. The main focus is on the FDA implementation by functional basis as described in section 4.2.3.

A.1 Basic Types of FDA Objects

There are three basic types of functional data objects, all are implemented as the `list` XploRe object:

`basisfd` defines a object containing functional basis (denoted by θ) in section 4.2.3. It is a `list` object containing following elements:

`fbname` - string, the name of the functional basis, supported are: `fourier, polynomial, bspline`, as described in section 4.2.3

`range` - (2×1) vector, `range[1]` min, `range[2]` max of interval J

`nbasis` - scalar, number of basis functions (L)

`param` - abstract set of elements, functional basis specific parameters that determine the functional basis `fbname`:

`fourier`, `param` is scalar, determining the length of period, $|J|$ in (3.12)

`polynomial`, `param` is scalar, determining shift – (ω) in (3.15), it is an optional parameter

`bspline`, `param` is $m \times 1$ vector, strictly non-decreasing knot sequence.

Note that if `nbasis` is even, `nbasis`=`nbasis`+1 will be used for `fourier` basis. Moreover, `nbasis`=m+ `order` -2 where `order` denotes the order of the splines for `bspline` basis.

`fdobject` defines an object containing functional data (denoted by $X_{\theta;i}, i = 1, \ldots, n$) in (4.15)

`basisfd` - object of type functional basis

`coef` - $n\times$ `nbasis` array of coefficients

`ldo` defines a linear differential operator (LDO denoted by $\mathcal{L}$), see (3.19) created by `createldo` quantlet see below, `ldo` is a list containing following elements:

`ldoname` name of xplore function (`proc-endp` environment) where the coefficients of linear differential differential operator are defined (a_1, ..., a_P) in (3.19), e.g. two LDOs: $a_1(t) = 1.25t$ and $a_2(t) = 0.5t^2$ are defined by function LDO1 (example taken from Ulbricht [2004]):

```
proc (wt) = LDO1 (evalarg)
wt = matrix (2, rows (evalarg),
     cols (evalarg))
wt[1,] = 1.25 .* evalarg
wt[2,] = 0.5 .* evalarg^2
endp
```

`derivs` $P \times 1$ vector defining order of derivatives in LDO.

A.1.1 Creating a FDA Data Object

The first step of the creation of the FDA data is the definition of basis and, optionally, definition of the LDO object, see section 3.3.4 for methodological issues. This can be done by `createfdbasis` and `createldo` respectively:

`fdbasis = createfdbasis (fbname,range,param)`
creates the fdbasis object on the interval `[range[1],range[2]]` using parameters `param`

`LDO = createldo (ldoname, derivs)`
creates the ldo object, using function (`ldoname`) and order of derivatives `derivs`

The input and output parameters are described above. More specific, for example, a 4th order bspline basis on interval $[0, 1]$ with the knot sequence $0.1, 0.2, \ldots, 1$ is created by:

```
fdbasis = createfdbasis
       ("bspline",0|1,12,(1:10)./10).
```

since `order=nbasis-m+2`. An LDO object $\mathcal{L} = a_1\mathcal{D} + a_2\mathcal{D}^3$ with a_1, a_2 defined above is created by:

```
ldo = createldo ("LDO1", 1|3).
```

The next step of the creation of the FDA data object, is the conversion of the discrete data into the (XploRe) functional data object (`fdobject`). This can be done using quantlet `data2fd`:

`fd = data2fd (y, argvals, basisfd, Lfd, W, lambda)`
converts an array `y` of function values, or penalized regression with `Lfd` and `lambda` observed on an array `argvals` of argument values into a functional data object

`y` - (T × n × r) array of discrete functional values. T is the number of grid points, n is the number of functions, r is the number of variables (A more general situation is handled here than that considered in section 4.2.3 where only one function at each point i is observed, this implies $r = 1$, see also (4.17)).

`argvals` - either a $(n \times T)$ matrix of argument values (corresponding to t_{ij} in (4.17)), or a $T \times 1$ matrix, in this case $t_j = t_{ij}$ for $i = 1, \ldots, n$ are assumed.

`basisfd` - an fdbasis object

`Lfd` - Linear differential operator. This can be a scalar (≥ 1) containing the order of derivative, e.g. `Lfd`=2 implies $\mathcal{L} = \mathcal{D}^2$. If an operator (3.19) with constant functions $a_p, p = 1, \ldots, P$ a $(P \times 2)$ is considered, then `Lfd` is a $(P \times 2)$ matrix, where the first column contains the coefficients, the second one the orders of derivatives. An operator with variable coefficients is implemented as a `ldo` object described above. The `Lfd` is an optional parameter.

`W` - Weighting matrix for weighted LS estimation. This is an optional parameter, the default value is the identity matrix (OLS estimation).

`lambda` - Parameter for roughness penalty (α in (3.18)). Optional parameter. If lambda is not specified it will be estimated by the data for each replication separately, see Ulbricht [2004] p. 60.

For an example of usage of `data2fd` - figure, displaying the input data (`y`) and approximated function (for one replication), see Figure 4.3.

Note that for the sake of compatibility between different statistical languages, these three objects correspond essentially to the `fd` and `basis` data classes of Ramsay [2003].

A.1.2 Statistical Procedures

`fdamean = fdamean(fdobject)`
creates a functional object with mean function from a functional object `fdobject`

The output of quantlet `fdamean` is a functional object.

The smoothed functional PCA is implemented in XploRe via the quantlet `fdaspca`, with the following syntax:

```
{fpcaresult,values,varprop,scores} =fdaspca(fdobject
                                                {lambda,
    lfd,npc,norm})
```
performs the smoothed functional PCA

The input parameters are:

`fdobject` list, functional object with n replications

`lambda` scalar, the smoothing parameter, default is 0

`npc` scalar, the number of (generalized) eigenfunctions, default is 4

`norm` string, normalization type, if norm="orthonorm" fpcaresult.coef are orthonormalized (with respect to the basis penalty matrix), if norm= "norm" the coefficients are renormalized to norm=1, default is "no" – no normalization is performed

The output variables are:

`fpcaresult` list, functional object

`values` `npc` $\times$ 1 vector, the (generalized) eigenvalues

`varprop` `npc` $\times$ 1 vector, the proportion of variance explained by the eigenfunctions

`scores` `N` $\times$ `npc` matrix, the principal components scores

For an example of usage of `fdaspca`, see Figure 4.7.

A.1.3 Evaluating and Visualizing FDA Data and its Functionals

The functional object can be evaluated on a certain grid, we may create a XploRe graphical object or directly create a plot using:

`evalmat = evalfd (evalarg, fd, Lfd)`
evaluates a functional data object `fd`, operated by a linear differential operator `Lfd`, at argument values in array `evalarg`

`grfd=grfd(fd,evalarg,Lfd,col,art,thick)`
creates a graphical object from the functional object `fd` operated by `Lfd`, using plotting grid `evalarg`. The line mask `col, art, thick` follow XploRe standards for `setmaskl`

`plotfd(fd,evalarg,Lfd,col,art,thick)`
plots the functional object `fd` operated by `Lfd`, using plotting grid `evalarg`. The line mask `col, art, thick` follow XploRe standards for `setmaskl`

`grfd=gr3dfd(fd,evalarg,Lfd,col,art,thick)`
creates a graphical object from the functional object `fd` operated by `Lfd`, using plotting grid `evalarg`. The line mask `col, art, thick` follow XploRe standards for `setmaskl`

`plot3dfd(fd,evalarg,Lfd,col,art,thick)`
plots the functional object `fd` operated by `Lfd`, using plotting grid `evalarg`. The line mask `col, art, thick` follow XploRe standards for `setmaskl`

`inprodmat = inprod(fdo1,fdo2{,Lfd1{,Lfd2{,JMAX{,EPS}}}})`

computes inner products based on fdobjects: `fdo1,fdo2` optionally transformed by operators `Lfd1, Lfd2` respectively. Maximum number of iterations for numerical integration is set by `JMAX`, default is 15, `EPS` - accuracy level, as determined by the extrapolation error estimate. The default value is 10^{-4}. The inner product is calculated by analytic or numerical integration, depending on the kind of input parameters.

The quantlet `evalarg` uses following important auxilirian quantlets:
1) `getbasismatrix` for evaluating the basis functions and `Fourierevalgd`,

`polyevalgd` and `Bsplineevalgd` evaluating the specific bases, see Ulbricht [2004] for details.

An example of `grfr` and `grfd3d` is plotted in figure 4.3.

For the variance, covariance and correlation functions we are faced with different situation because using functional basis expansion we need to evaluate a quadratic or bilinear form, which is not directly supported by the existing XploRe functional data objects. For this reason we designed quantlets for evaluation, creation of graphical objects and plotting of this functional data characteristics. The syntax of these quantlets is the following:

Variance function (one dimensional):

`fdvar=evalfdavar(evalarg,fdobject)`
evaluates the variance function of a functional data object `fd`, at argument values in an array `evalarg`

`gdfvar=grfdavar(fd,evalarg,col,art,thick)`
creates a graphical object from the variance function of a functional object `fd`. The line mask `col, art, thick` follow XploRe standards for `setmaskl`

`plotfdavar(fd,evalarg,col,art,thick)`
plots the variance function of a functional object `fd`. The line mask `col, art, thick` follow XploRe standards for `setmaskl`

Covariance and Correlation function (two dimensional functions) - surfaces:

`fdcov=evalfdacov(evalarg,fdobject)`
evaluates the cov function of a `fdobject` at the vector `evalarg` $\times$ `evalarg`

`fdcov=evalfdacorr(evalarg,fdobject)`
evaluates the corr function of a `fdobject` at the vector `evalarg` $\times$ `evalarg`

`gs=grfdacov(fdobject,evalarg,col)`
creates a graphical object with cov surface, using plotting grid `evalarg` × `evalarg` and color `col`

`gs=grfdacorr(fdobject,evalarg,col)`
creates a graphical object with corr surface, using plotting grid `evalarg` × `evalarg` and color `col`

`plotfdacov(fdobject,evalarg,col,plot3Ds)`
plots the cov surface of `fdobject`, using plotting grid `evalarg` × `evalarg` and color `col`, if `plot3Ds`="3D" the quantlet `plot3d` will be used for plotting

`plotfdacorr(fdobject,evalarg,col,plot3Ds)`
plots the corr surface of `fdobject`, using plotting grid `evalarg` × `evalarg` and color `col`, if `plot3Ds`="3D" the quantlet `plot3d` will be used for plotting

`plotboxfd(fd,evalarg,Lfd,wfac)`
plots the pointwise boxplots of `fd` after applying linear differential operator `Lfd` at `evalarg`, the length of the whiskers is controlled by `wfac`

The input parameters for the quantlets above are:

`fd,fdobject`, lists, XploRe functional objects

`evalarg`, $m \times 1$ vector, grid for evaluation

`Lfd` - Linear differential operator. This can be a scalar (≥ 1) containing the order of derivative, e.g. `Lfd`=2 implies $\mathcal{L} = \mathcal{D}^2$. If an operator (3.19) with constant functions $a_p, p = 1, \ldots, P$ a $(P \times 2)$ is considered, then `Lfd` is a $(P \times 2)$ matrix, where the first column contains the coefficients, the second one the orders of derivatives. An operator with variable coefficients is implemented as a `ldo` object described above. The `Lfd` is an optional parameter. For graphical and plotting quantlets `Lfd` as

vector is not supported, in this case Ldf[1] will be used and a warning will be returned.

`col` vector of integers, color, parameter for `setmaskl`, default is 0

`art` vector of integers, art type, parameter for `setmaskl`, default is 1

`thick` vector of integers, thickness type, parameter for `setmaskl`, default is 2

`plot3Ds` string, if string is = "3D" plot3D quantlet will be used for plotting

The functions $Cov, Corr$ are evaluated at the points $t_i, s_j \in$ evalarg.

Example of `grfdacov`, `grfdacorr` and `plotboxfd` see Figure 4.5, Figure 4.5 and figure 4.4 respectively.

A.1.4 Further Comments

As already mentioned, the "low level" (object types and commands) parts of the FDA library are more or less compatible with the packages for R by Ramsay [2003]. In fact the R package is older and more general, the developing of this library was motivated by that work. The advantage is that the developer of the further functional data procedures can easily adapt their work from one environment to another. The graphical and plotting functions however are motivated more by "XploRe-philosophy". One basic difference at the present version of the XploRe FDA library are the multivariate FDA objects: (e.g. at the same time we observe a vector and not just one number). This situation is supported currently only at the object and low level quantlets as `evalfd` and `inprod`. The statistical, graphical and plotting routines are not directly supporting this, in case of multivariate objects, only the first functional value is taken into account and a warning is returned. Extraction of the univariate fdobject from a multivariate can be done via quantlet `extractfd`. Future generalization is easily possible.

```
fdobjectout = extractfd(fdobject,varind,repind)
      extracts specified variable or replications from a FDOobject
```

The input and output parameters are:

`fdobjectout,fdobject`, lists, XploRe functional objects

`varind`, vector or scalar, index of variable to be extracted

`repind`, vector or scalar, index of replications to be extracted

Note, that the FDA analysis consists of analyzing a sample of functions. In the first step, we are usually faced with "sample of regression problems". It is known that a lot of effort was spent on the automatization of the estimation techniques in the regression however it is also known that none of the technique is generally always working perfectly. Clearly a error done at this stage, namely by representing discrete data as functions, may have a crucial effect by the interpretation of results of further studies. An interactive tool that allows e.g. visualization and analysis of individual function, their functionals and corresponding discrete data, scrolling between them, might be very useful.

Unfortunately here we are reaching the limits of the visualization and interaction potential of standard statistical packages as R or XploRe. Hence an important stream of further research on this area could be the development of the specialized add-ons or specialized software packages. An example of this direction of research can be found in Jank et al. [2006] with additional bibliographic remarks.

Appendix B

Figures – Phenomenology of IVs

The first part of this section presents the additional plots and results of the static descriptive analysis for the IV functions and surfaces discussed in chapter 3.

Figure B.1 plots the standard deviation as a function of time-to-maturity for the smoothed IVS for ATM options on the yearly basis. The figures B.2 up to figure B.3 show yearly averaged IVS, for interpretation and discussion see section 3.7.

Figures B.4–B.7 visualize the design densities of the IVS as functions of (futures) moneyness and time-to-maturity, for discussion see section 3.6.1.

Last figure of this section, figure B.8 plots the factor functions for two-dimensional factor model (moneyness and time-to-maturity) for the IVS, as proposed by Detlefsen and Härdle [2006], see section 4.5.1 for details.

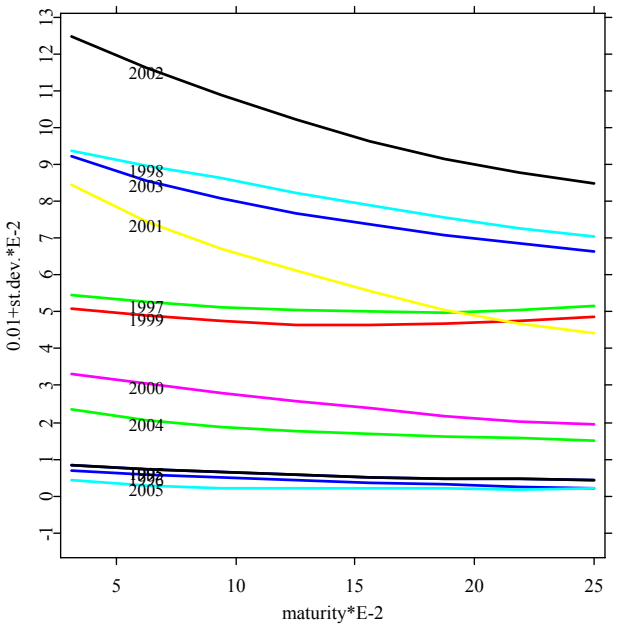

Figure B.1: Standard deviation for smoothed implied volatility surfaces (by local linear estimate) for ATM for years 1995 up to 2005 (for 2005 only data up to June were available)

.

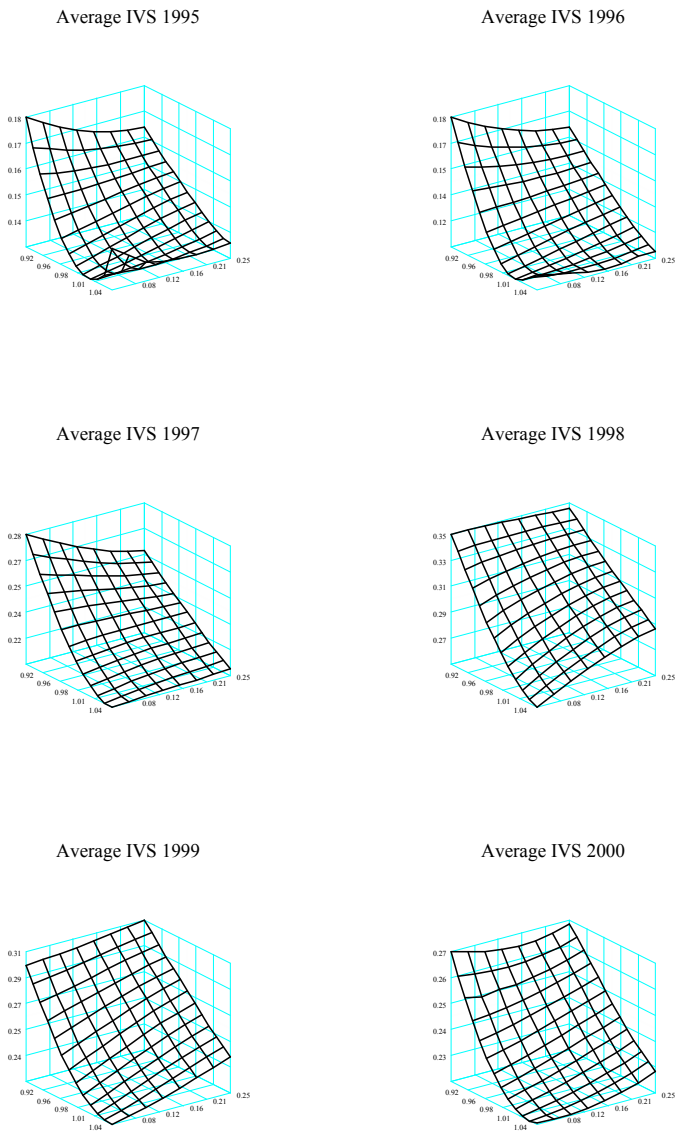

Figure B.2: Yearly averaged smoothed (by local linear estimate) IVs surfaces for period from 1995 to 2000.

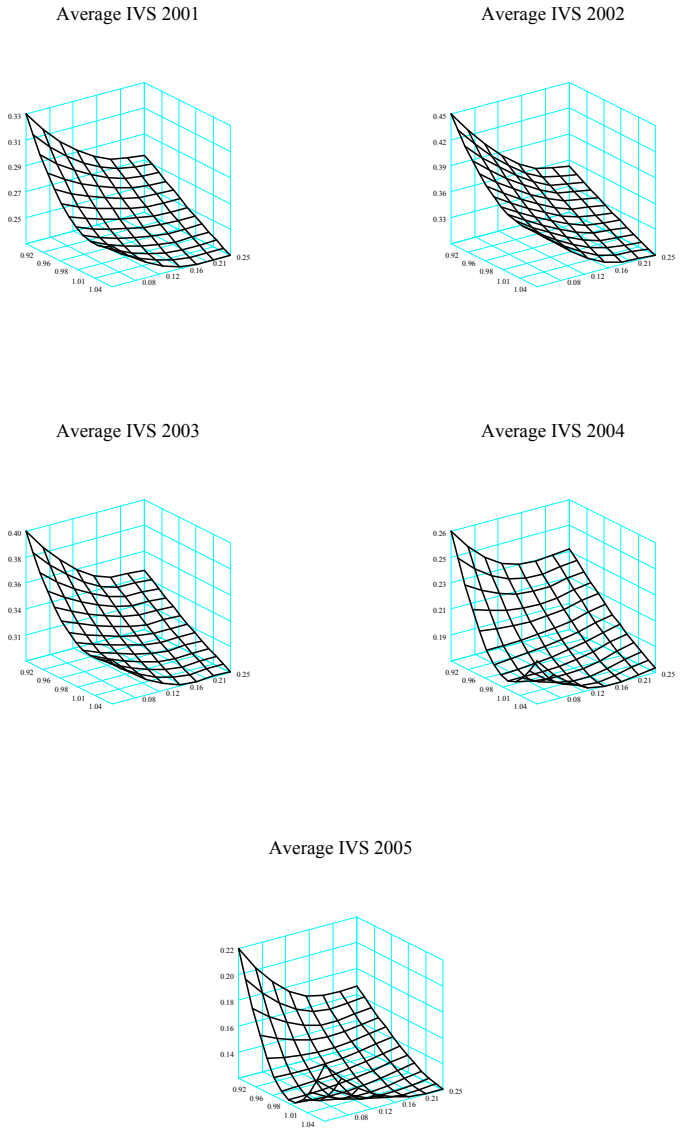

Figure B.3: Yearly averaged smoothed (by local linear estimate) IVs surfaces for period from 2001 to 2005 (for 2005 only data up to June were available)

.

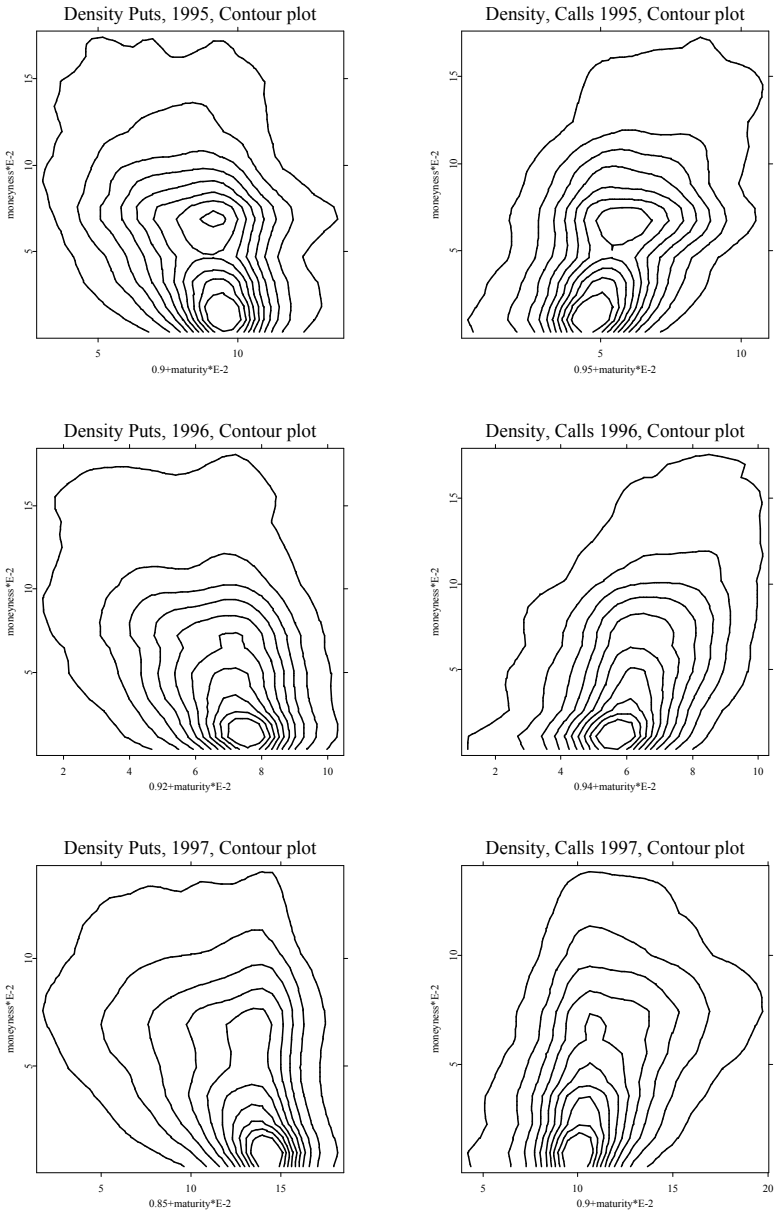

Figure B.4: Contour plots for liquidity densities, Puts (left), Calls (right) for years 1995 – 1997

.

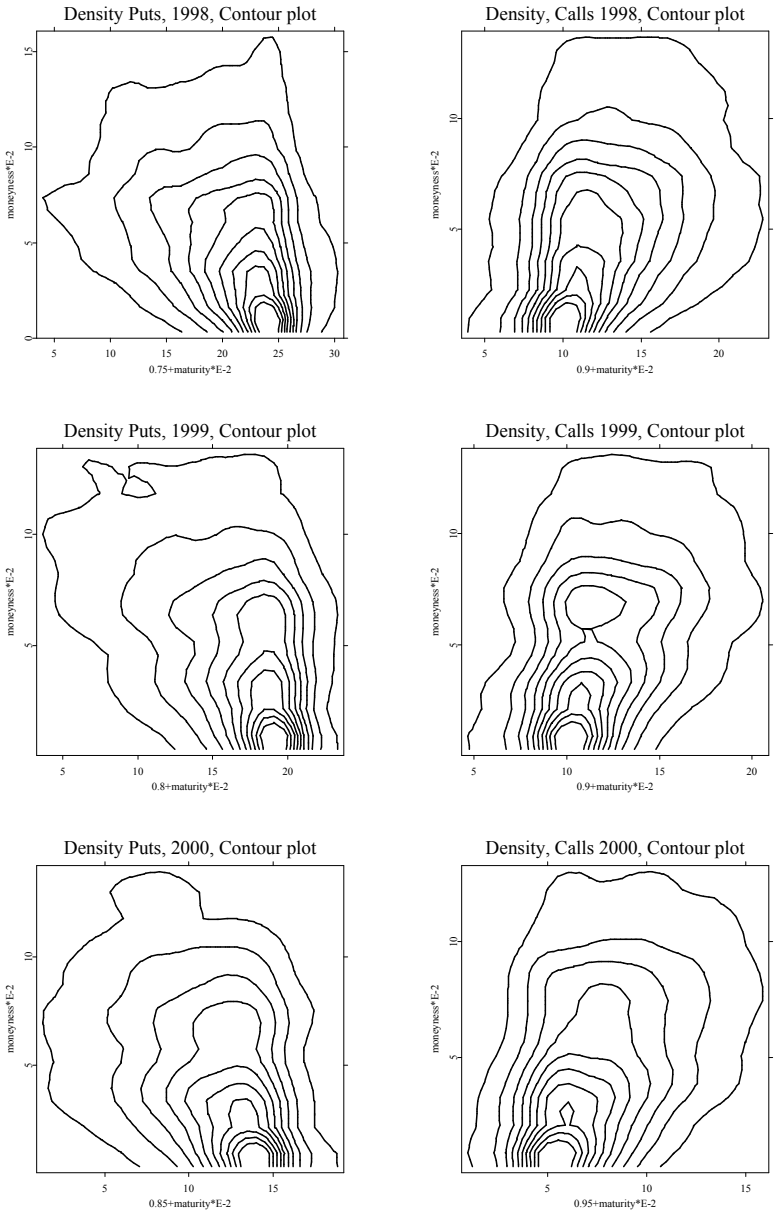

Figure B.5: Contour plots for liquidity densities, Puts (left), Calls (right) for years 1998 – 2000

.

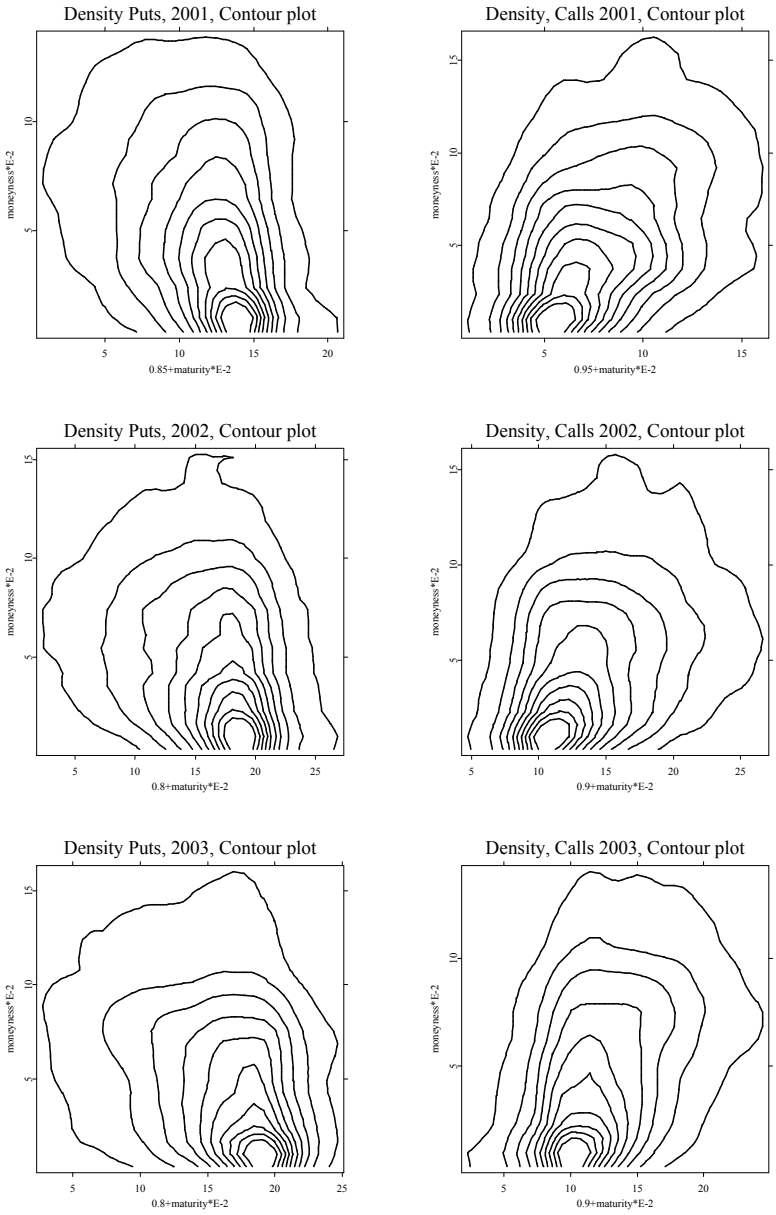

Figure B.6: Contour plots for liquidity densities, Puts (left), Calls (right) for years 2001 – 2003

.

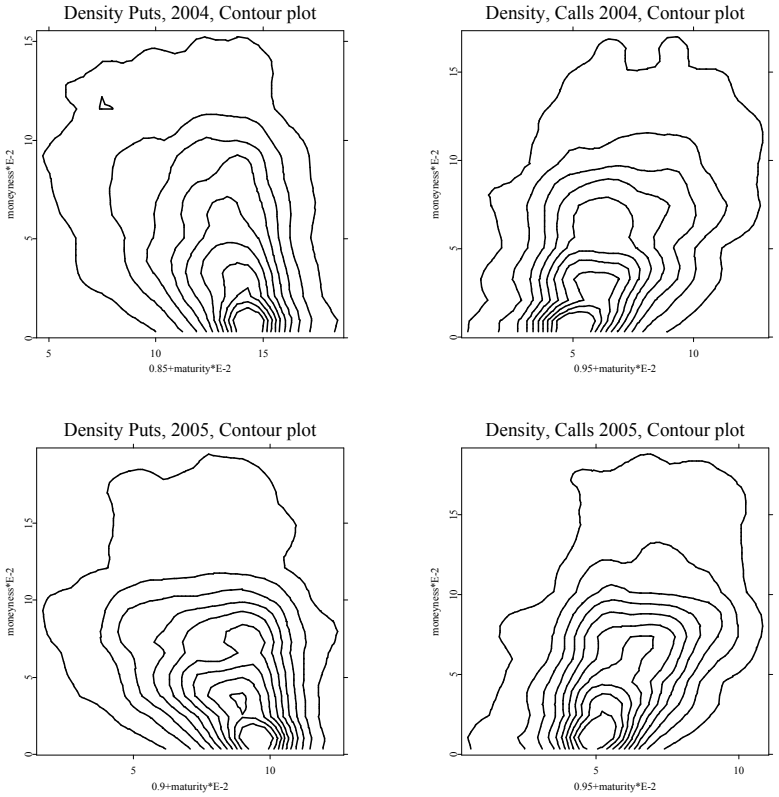

Figure B.7: Contour plots for liquidity densities, Puts (left), Calls (right) for years 2004 – 2005 (for 2005 only data up to June were available)

.

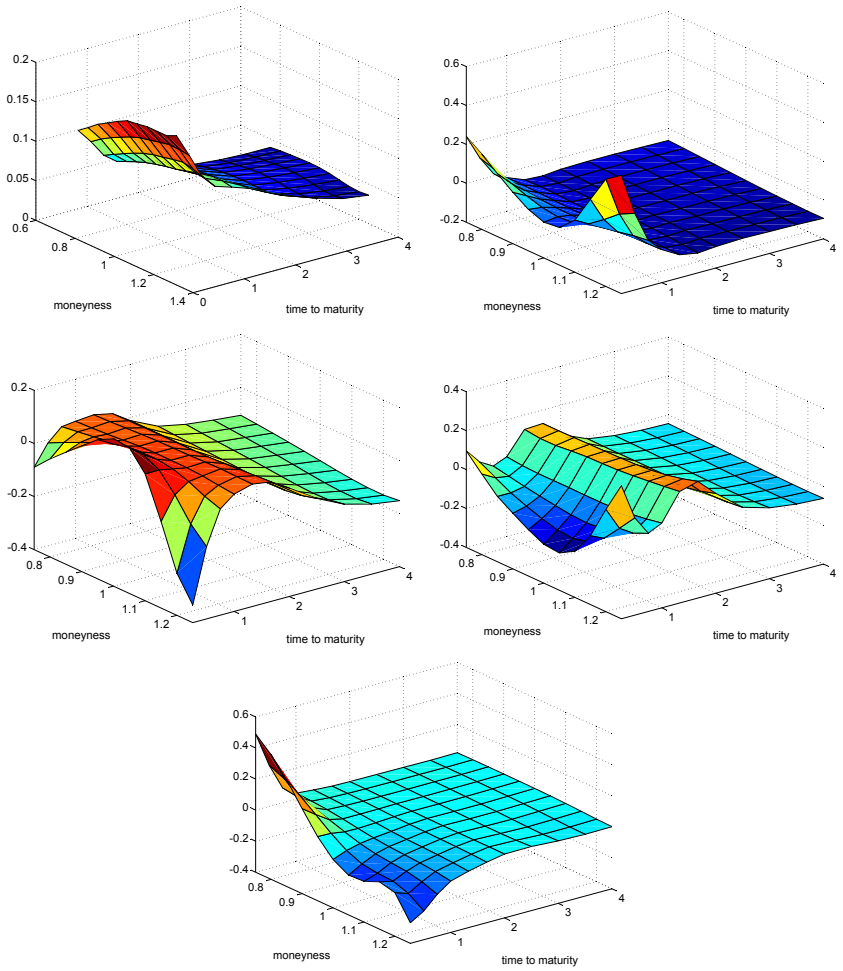

Figure B.8: Factor functions based on the two-dimensional PCA analysis on IVS of ODAX, as proposed by Detlefsen & Härdle (2006)

.

B.1 SPD as a Function of IV

First remember BS formula (2.6):

$$C_t^{BS}(S_t, K, \tau, r, \sigma) = S_t\Phi(d_1) - Ke^{-r\tau}\Phi(d_2) \tag{B.1}$$

where

$$d_1 = \frac{\ln(S_t/K) + (r + 1/2\sigma^2)\tau}{\sigma\sqrt{\tau}} \tag{B.2}$$

$$d_2 = d_1 - \sigma\sqrt{\tau} = \frac{\ln(S_t/K) + (r - 1/2\sigma^2)\tau}{\sigma\sqrt{\tau}}. \tag{B.3}$$

Note that in B.1 we consider σ as IV, i.e. sigma for witch BS price equals the observed price and as a function of K and τ. Using some simple algebra:

$$\phi(d_1) = \phi(d_2)\frac{K}{S_t}e^{-r\tau}. \tag{B.4}$$

Using standard rules for derivaties we obtain from (B.2) and (B.3):

$$\frac{\partial d_1}{\partial K} = \frac{-\frac{1}{K} + \frac{\partial\sigma}{\partial K}\sigma\tau - d_1\frac{\partial\sigma}{\partial K}\sqrt{\tau}}{\sigma\sqrt{\tau}} \tag{B.5}$$

$$\frac{\partial d_2}{\partial K} = \frac{-\frac{1}{K} - \frac{\partial\sigma}{\partial K}\sigma\tau - d_2\frac{\partial\sigma}{\partial K}\sqrt{\tau}}{\sigma\sqrt{\tau}} \tag{B.6}$$

From (B.5) and (B.6) follows also:

$$\frac{\partial d_1}{\partial K} - \frac{\partial d_2}{\partial K} = \frac{\partial\sigma}{\partial K}\sqrt{\tau}. \tag{B.7}$$

From (B.1) we obtain

$$\frac{\partial C_t^{BS}}{\partial K} = S_t\phi(d_1)\frac{\partial d_1}{\partial K} - Ke^{-r\tau}\phi(d_2)\frac{\partial d_2}{\partial K} - e^{-r\tau}\Phi(d_2) \tag{B.8}$$

(B.8) together with (B.4) and (B.7) yields

$$\begin{aligned}\frac{\partial C_t^{BS}}{\partial K} &= S_t\phi(d_1)\left[\frac{\partial d_1}{\partial K} - \frac{\partial d_2}{\partial K}\right] - e^{-r\tau}\Phi(d_2) \\ &= S_t\phi(d_1)\frac{\partial\sigma}{\partial K}\sqrt{\tau} - e^{-r\tau}\Phi(d_2)\end{aligned} \tag{B.9}$$

(B.9), the fact that $\frac{\partial\phi(d_1)}{\partial K} = -d_1\phi(d_1)\frac{\partial d_1}{\partial}$ together with (B.4), (B.5) and (B.6) lead to

$$\begin{aligned}
\frac{\partial^2 C_t^{BS}}{\partial K^2} &= S_t\phi(d_1)\sqrt{\tau}\left[-d_1\frac{\partial d_1}{\partial K}\frac{\partial\sigma}{\partial K} + \frac{\partial^2\sigma}{\partial K^2} - \frac{1}{K\sqrt{\tau}}\frac{\partial d_2}{\partial K}\right] \\
&= S_t\phi(d_1)\sqrt{\tau}\left\{\frac{1}{K^2\sigma\tau} + \frac{\partial\sigma}{\partial K}\frac{2d_1}{\sqrt{\tau}K\sigma} + \frac{\partial^2\sigma}{\partial K^2} + \left(\frac{\partial\sigma}{\partial K}\right)^2\left[\frac{d_1^2}{\sigma} - d_1\sqrt{\tau}\right]\right\} \\
&= S_t\phi(d_1)\sqrt{\tau}\left\{\frac{1}{K^2\sigma\tau} + \frac{\partial\sigma}{\partial K}\frac{2d_1}{\sqrt{\tau}K\sigma} + \frac{\partial^2\sigma}{\partial K^2} + \left(\frac{\partial\sigma}{\partial K}\right)^2\left[\frac{d_1 d_2}{\sigma}\right]\right\} \quad \text{(B.10)}
\end{aligned}$$

Finally from (B.10) we can express SPD as:

$$\begin{aligned}
q_{t,\tau} &= e^{r\tau}\frac{\partial^2 C_t(K,T)}{\partial K^2} \\
&= e^{r\tau}S_t\phi(d_1)\sqrt{\tau}\left\{\frac{1}{K^2\sigma\tau} + \frac{\partial\sigma}{\partial K}\frac{2d_1}{\sqrt{\tau}K\sigma} + \frac{\partial^2\sigma}{\partial K^2} + \left(\frac{\partial\sigma}{\partial K}\right)^2\left[\frac{d_1 d_2}{\sigma}\right]\right\} \\
&= F_t\phi(d_1)\sqrt{\tau}\left\{\frac{1}{K^2\sigma\tau} + \frac{\partial\sigma}{\partial K}\frac{2d_1}{\sqrt{\tau}K\sigma} + \frac{\partial^2\sigma}{\partial K^2} + \left(\frac{\partial\sigma}{\partial K}\right)^2\left[\frac{d_1 d_2}{\sigma}\right]\right\} \quad \text{(B.11)}
\end{aligned}$$

Formula (B.11) is of the form (slightly corrected) in Fengler [2005b]. Using (B.4) we can rewrite the last formula into the form presented (without derivation) in Brunner and Hafner [2003]:

$$q_{t,\tau}(K) = \phi(d_2)\left\{\frac{1}{K\sigma\sqrt{\tau}} + \frac{\partial\sigma}{\partial K}\frac{2d_1}{\sigma} + \frac{\partial^2\sigma}{\partial K^2}\sqrt{\tau}K + \left(\frac{\partial\sigma}{\partial K}\right)^2\frac{d_1 d_2}{\sigma}\sqrt{\tau}K.\right\}$$

More over the formula (B.8) gives also the form of the distribution function connected with the SPD:

$$\begin{aligned}
Q_{t,S_T}(K) &= e^{r\tau}\frac{\partial C_t(K,T)}{\partial K} \\
&= F_t\phi(d_1)\frac{\partial d_1}{\partial K} - K\phi(d_2)\frac{\partial d_2}{\partial K} - \Phi(d_2). \quad \text{(B.12)}
\end{aligned}$$

Bibliography

C. Alexander. Common correlation and calibrating the lognormal forward rate model. Discussion Papers in Finance 2002-18, ISMA Centre, University of Reading, 2002.

A. Andersen and T. Wagener. Extracting risk neutral probability densities by fitting implied volatility smiles: some methodological points and an application to the 3m euribor futures option prices. *Working Paper, European Central Bank*, No 2002-198, 2002.

A. Andriyashin, M. Benko, W. Härdle, R. Timofeev, and U. Ziegenhagen. Colour harmonization in car manufacturing processes. *Applied Stochastic Models in Business and Industry*, 22, 2006.

A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, 1980.

K.J. Arrow. The role of securities in the optimal allocation of risk-bearing. *Review of Economical Studies*, 31(2):91–96, 1964.

D. S. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9:69–107, 1996.

M. Benko. Selected problems in nonparametric regression. Master thesis, Charles University, 2002. Prague, in Slovak/Czech (Vybrané problémy neparametrické regrese).

M. Benko. Functional principal components analysis, implementation and applications. Master thesis, Department of Business and Economics, 2004. Humboldt-Universität zu Berlin.

M. Benko and W. Härdle. Common functional implied volatility analysis. In P. Čížek, W. Härdle, and R. Weron, editors, *Statistical Tools in Finance*. Springer-Verlag, Berlin, Heidelberg, 2005.

M. Benko, M. Fengler, W. Härdle, and M. Kopa. On extracting information implied in options. *Submitted to Computational Statistics*, 2006a.

M. Benko, W. Härdle, and A. Kneip. Common functional principal components. *Submitted to Annals of Statistics*, 2006b.

D. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, 1999.

P. Besse and J. Ramsay. Principal components of sampled functions. *Psychometrika*, 51:285–311, 1986.

F. Black. Living up to the model. In P. Field and R. Jaycobs, editors, *From Black-Scholes to Black Holes: New Frontiers in Option Pricing*, pages 17–20. Risk Magazine Ltd, London, 1992.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

O. Blaskowitz, H. Herwartz, and G. de Cadenas Santiago. Modeling the fibor/euribor swap term structure: An empirical approach. Discussion paper, SfB 649, Humboldt-Universität zu Berlin, 2005.

R. Bliss and N. Panigirtzoglou. Testing the stability of implied probability density functions. *Journal of Banking and Finance*, 26:381–422, 2002.

Deutsche Börse. *Leitfaden zu den Volatilitärsindizes der Deutschen Börse*. Deutsche Börse, Frankfurt am Main, 2006a.

Deutsche Börse. *VDAX-NEW: Der neue Volatilitätsindex der Deutschen Börse*. Deutsche Börse AG, 2006b.

D. Breeden and R. Litzenberger. Price of state-contingent claims implicit in options prices. *Journal of Business*, 51:621–651, 1978.

M. Britten-Jones and A. J. Neuberger. Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, 55(2):839–866, 2000.

O. Brockhaus, M. Farkas, A. Ferraris, D. Long, and M. Overhaus. *Equity derivatives and market risk models*. Risk Books, London, 2000.

B. Brunner and R. Hafner. Arbitrage-free estimation of the risk-neutral density from the implied volatility smile. *Journal of Computational Finance*, 7(1):75–106, 2003.

R. Cont and J. da Fonseca. The dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60, 2002.

J. E. Cox and S. A. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 76:145–166, 1976.

J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12:136–154, 1982.

G. Debreu. *Theory of Value*. Wiley, 1956.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300:463–520, 1994.

K. Detlefsen and W. Härdle. Dynamics of implied volatility surfaces revisited. Discussion paper, SfB 649, Humboldt-Universität zu Berlin, 2006.

F. X. Diebold and C. Li. Modeling and forecasting the term structure of government bond yields. *Journal of Econometrics*, 2005. Forthcoming.

J. Dupačová, J. Hurt., and J. Štěpán. *Stochastic Modeling in Economics and Finance*. Kluwer Academic Publishers, 2002.

B. Dupire. Pricing with a smile. *RISK*, 7(1):18–20, 1994.

J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London, 1996.

J. Fan and L. Huang. Nonparametric estimation of quadratic regression functionals. *Bernoulli*, 5:927–949, 1999.

M. R. Fengler. Arbitrage-free smoothing of the implied volatility surface. Discussion paper, SfB 649, Humboldt-Universität zu Berlin, 2005a.

M. R. Fengler. *Semiparametric Modeling of Implied Volatility*. Lecture Notes in Finance. Springer-Verlag, Berlin, Heidelberg, 2005b.

M. R. Fengler, W. Härdle, and P. Schmidt. Common factors governing VDAX movements and the maximum loss. *Journal of Financial Markets and Portfolio Management*, 16(1):16–29, 2002.

M. R. Fengler, W. Härdle, and C. Villa. The dynamics of implied volatilities: A common principle components approach. *Review of Derivatives Research*, 6:179–202, 2003.

M. R. Fengler, W. Härdle, and E. Mammen. A dynamic semiparametric factor model for implied volatility string dynamics. Discussion Paper 20, SfB 649, Humboldt-Universität zu Berlin, 2005.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis. Theory and Practice*. Springer-Verlag, Berlin, Heidelberg, 2006.

F. Ferraty, P. Vieu, and S. Viquier-Pla. A note on factor-based comparison of groups of curves. *Elsevier Science*, 2006. to appear.

B. Flury. *Common Principal Components and Related Multivariate Models*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Son, New York, 1988.

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. Wiley Series in Probability and Mathematical Statistics. Walter de Gruyter, Berlin, New York, 2002.

I.I. Gihman and A.V Skorohod. *The theory of stochastic process II*. Springer-Verlag, Berlin, Heidelberg, 1973.

R. Hafner. *Stochastic Implied Volatility*. Springer-Verlag, Berlin, Heidelberg, 2004.

R. Hafner and M. Wallmeier. The dynamics of DAX implied volatilities. *International Quarterly Journal of Finance*, 1(1):1–27, 2001.

P. Hall, J.W. Kay, and D.M. Titterington. Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77: 520–528, 1990.

W. Härdle. Robust regression function estimation. *Journal of Multivariate Statistics*, 12:624–635, 1984.

W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, UK, 1990.

W. Härdle and Z. Hlávka. Dynamics of state price densities. Discussion Paper 2005-21, SfB 649, Humboldt-Universität zu Berlin, 2005.

W. Härdle and L. Simar. *Applied Multivariate Statistical Analysis*. Springer-Verlag, Berlin, Heidelberg, 2003.

W. Härdle, S. Klinke, and M. Müller. *XploRe Learning Guide*. Springer-Verlag, Berlin, Heidelberg, 2000.

W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer-Verlag, Berlin, Heidelberg, 2004.

J. Harrison and D. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

J.D. Hart. *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer-Verlag, Berlin, Heidelberg, 1997.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, Berlin, Heidelberg, 2002.

L. Hentschel. Errors in implied volatility estimation. *Journal of Financial and Quantitative Analysis*, 38:779–810, 2003.

S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

C.J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Jounal of finance*, 42:281–300, 1987.

J. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, New Jersey, USA, 6th edition, 2003.

K.M.S. HUMAK. *Statistische Methoden der Modellbildung Band I*. Akademie Verlag, Berlin, 1997.

W. Jank, G. Shmueli, C. Plaisant, and B. Shneiderman. Visualizing functional data with an application ti ebay's online auctions. In C. Chen, W. Härdle, and A. Unwin, editors, *Handbook of Data Visualization*. Springer-Verlag, 2006. to appear.

G.J. Jiang and Y.S. Tian. Model-free implied volatility and its information content. *Review of Financial Studies*, 18:1305–1342, 2003.

G. G. Judge, W. E. Griffiths, R. C. Hill, and H. Lütkepohl. *Introduction to the Theory and Practice of Econometrics*. John Wiley and Sons, New York, 1988.

N. Kahalé. An arbitrage-free interpolation of volatilities. *RISK*, 17(5):102–106, 2004.

A. Kneip and K. Utikal. Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, 96(454):519–531, 2001.

P.J. Knez, R. Litterman, and J. Scheinkman. Explorations into factor explaining money market returns. *Journal of Finance*, 49:1861–1882, 1994.

H. A. Latané and J. Rendelman. Standard deviations of stock price ratios implied in option prices. *Journal of Finance*, 31:369–381, 1976.

J. Lukeš and J. Malý. *Measure and Integral*. Matfyz Press, Prague, 1995.

R. C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(Spring):141–183, 1973.

R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

M. Musiela and M. Rutkowski. *Martingale Methods for Financial Modelling*. Springer-Verlag, Berlin, Heidelberg, 1997.

S. Natanberg. *Option Volatility and Pricing*. Probus Publishing, Cambridge, 1994.

C. R. Nelson and A. F. Siegel. Parsimonious modeling of yield curves. *Journal of Business*, 60:473–489, 1987.

M. Pešta. Isotonic regression in sobolev spaces. Master thesis, Charles University, 2006. Prague.

S. D. Pezzulli and B. W. Silverman. Some properties of smoothed principal components analysis for functional data. *Computational Statistics*, 8:1–16, 1993.

J. Ramsay. Matlab, r and s-plus functions for functional data analysis. Technical report, McGill University, 2003. ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns.

J. Ramsay and B. W. Silverman. *Applied Functional Data Analysis*. Springer-Verlag, Berlin, Heidelberg, 2002.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer-Verlag, Berlin, Heidelberg, 2nd edition, 2005.

R. Rebonato. *Volatility and Correlation*. John Wiley & Sons Ltd, 1999.

E. Renault. Econometric models of option pricing errors. In D. M. Kreps and K. F. Wallis, editors, *Advances in Economics and Econometrics, Seventh World Congress*, Econometric Society Monographs, pages 223–278. Cambridge University Press, Cambridge, 1997.

J. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of Royal Statistical Society, Ser. B*, 53:233–243, 1991.

J. Rosenberg. Implied volatility functions: A reprise. *Journal of Derivatives*, 7:51–64, 2000.

D. Ruppert, M.P. Wand, and Carroll. *Semiparametric Regression*. Cambridge University Press, Cambridge, 2003.

P. A. Samuelson. Rational theory of warrant prices. *Industustrial Management Review*, 6:13–31, 1965.

D. W. Scott. *Multivariate Density Estimation*. John Wiley and Sons, New York, 1992.

D. Shimko. Bounds on probability. *RISK*, 6(4):33–37, 1993.

S. E. Shreve. *Stochastic Calculus for Finance II, Continuous-Time Models*. Springer-Verlag, Berlin, Heidelberg, 2004.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman Hall, London, 1990.

B. W. Silverman. Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, 24:1–24, 1996.

V. Spokoiny. *Local Parametric Methods in Nonparametric Estimation*. Springer-Verlag, Berlin, Heidelberg, 2006.

R.G. Tompkins. *Options Analysis*. Probus Publishing, Cambridge, 1994.

J. Ulbricht. Repressenting functional data as smooth functions. Master thesis, Department of Business and Economics, 2004. Humboldt-Universität zu Berlin.

S. Viguier-Pla. Factor-based comparision of k populations. *Statistics*, 38(1): 1–15, 2004.

R. Zagst. *Interest Rate Management*. Springer-Verlag, Berlin, Heidelberg, 2002.

Frequently Used Notation

Abbreviation	Explanation
ATM	at-the-money
BS	Black-Scholes
DAX	German stock index, (registered trademark)
EUREX	European derivative exchange, operated by Deutsche Börse AG and SWX Swiss exchange (registered trademarks)
EURIBOR	Euro interbank offered rate
VDAX	DAX volatility index (registered trademark)
FDAX	futures on DAX (registered trademark)
IV	implied volatility
IVS	implied volatility surface
ITM	in-the-money
ODAX	option on DAX (registered trademark)
OTC	over-the-counter
OTM	out-of-the-money
SPD	state-price (risk-neutral) density
S_t	price process of the (underlying) asset
C_t	price of an European call
P_t	price of an European call
$F_{t,\tau}$	price of a future
σ	(implied) volatility

Abbreviation	Explanation
$\mathbf{A}^\top$	transpose of a matrix $\mathbf{A}$
$\det(\mathbf{A})$	determinant of a square matrix $\mathbf{A}$
$\mathcal{D}$	differential operator ($\mathcal{D}^p$ denotes the pth derivative)
$\boldsymbol{I}$	indicator function
$\mathbb{N}^d$	set of the d-dimensional vectors of natural numbers
$\varnothing$	Landau 'o', $a_n = \varnothing(b_n)$ denotes $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$
$\mathcal{O}$	Landau 'O', $a_n = \mathcal{O}(b_n)$ denotes $\lim_{n\to\infty} \frac{a_n}{b_n} = c$, $0 < c < \infty$.
$\varnothing_P$	$a_n = \varnothing_P(b_n)$ denotes $\forall \varepsilon > 0$, $\lim_{n\to\infty} P(\lvert\frac{a_n}{b_n}\rvert > \varepsilon) = 0$
$\mathcal{O}_P$	$a_n = \mathcal{O}_P(b_n)$ denotes $\forall \varepsilon > 0$, $\exists N, M$, such that $P(\lvert\frac{a_n}{b_n}\rvert > M) < \varepsilon, \forall n > N$
$\mathbb{R}^d$	d dimensional real valued vector space
$\langle a, b\rangle$	scalar product of a, b – elements of some vector space
$\lvert\lvert a\rvert\rvert$	norm of a – a is an element of some vector space
$\stackrel{\text{def}}{=}$	is defined as
$\mathcal{H}^*$	dual space to the vector space $\mathcal{H}$
$\text{Ker}(\mathcal{L})$	kernel space of an operator $\mathcal{L}$
$\text{supp}(X)$	support of a function X
$[u, l]$	closed interval
(u, l)	open interval
Abbreviation	Explanation
a.s.	almost sure
c.d.f.	cumulative distribution function
ϕ	p.d.f. of a standard normal random variable
Φ	c.d.f. of a standard normal random variable
E	expected value
FDA	functional data analysis
FPC(A)	functional principal components (analysis)
$U[l, u]$	uniform distribution on an interval $[l, u]$
$N(\mu, \Sigma)$	normal distribution with mean μ and covariance Σ
NW	Nadaraya Watson (estimate)
(A)MSE	(As.) Mean Square Error $MSE(\hat{X}(t)) = E[\hat{X}(t) - X(t)]^2$
(A)MISE	(As.) mean integrated square error $MISE(\hat{X}) = \int MSE(\hat{X}(t))dt$
p.d.f.	probability density function
PCA	principal components analysis

Acknowledgement

First, I would like to thank to Professor Wolfgang Härdle for supervising and supporting me through the whole time of my Ph.D. studies, he introduced me to the world of nonparametrics and encouraged me to work on the analysis of functional data and empirical finance. His almost limitless energy was motivating me day-by-day.

The theoretical part of the thesis is based on the results of close cooperation with Professor Alois Kneip, whose extraordinary deep knowledge and experience in nonparametrics and functional data analysis helped me a lot in understanding of these methods and many discussions with him prevented me from roundabout ways, I'm also highly appreciating his hospitality during my several visits at the University of Mainz.

I owe much to many colleagues and researchers for sharing their time with me by numberless discussions and consultations during my work, among other this were: Professor James Ramsay, Professor Bern Rönz, Professor Vladimir Spokoiny, Szymon Borak, Kay Detlefsen, Matthias Fengler, Zdeněk Hlávka, Jan Ulbricht, Miloš Kopa, Julius Mungo, Enzo Giacomini, Roman Timofeev, Philippe Vieu as well as all other members of the STAPH group at University Toulouse and of course my thanks goes to all members of the Institute for Statistics at Humboldt University, C.A.S.E. and SFB 649 for friendly atmosphere and encouragement.

I gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft by the Sonderforschungsbereich 649 *Ökonomisches Risiko.*

Last but certainly not least I'm deeply indebted to my family for having been there whenever I needed them, especially the constant support and patient tolerance of my wife Lenka made the finishing of this thesis possible.

Berlin, October 22, 2006

Michal Benko

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzte Hilfsmittel, über dir mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.